

Chinese hesitancy on avian flu

The time has come for China to start pulling its weight as a participant in the global response to bird flu — and to learn to collaborate more openly.

Beijing is currently undertaking some spectacularly expensive preparations for hosting the 2008 Olympics, and its government officials talk of sending a man to the Moon. But when it comes to the more prosaic matter of taking elementary precautions against the palpable economic and public-health risks posed by bird flu, the world's most ebullient industrial power is pleading poverty.

When representatives of about 100 countries met in the Chinese capital last week to discuss funding for countermeasures to the spread of avian influenza in nations far poorer than China itself, the host country was ready to offer only US\$10 million of the \$1.9 billion pledged in total over three years. China can point to the resource requirements of its own mitigation strategies as a reason for its reluctance to send money abroad. But this explanation would carry more weight if the country had put convincing strategies in place and openly shared information and biological samples with the outside world. So far, it has failed to do so.

Epidemiologists say that the most important initial steps in constraining both the spread of avian flu and its mutation into a strain that could cause a global human pandemic are the establishment of monitoring systems to track the virus in poultry and wild birds, and the rapid diagnosis and follow-up of all human cases. Needy countries at high risk of exposure to the virus, such as Indonesia and Cambodia, urgently require help in setting up such systems.

China has some 14 billion poultry, many of them scattered about in backyard farms in intimate contact with other livestock and people, so it is one of the most likely breeding-grounds for a pandemic virus. Collating accurate information about the virus populations circulating in these birds is a considerable challenge. Outsiders like to accuse Chinese government officials of cover-ups, but the reality is often that the officials themselves don't know what is happening on the ground.

Every nation struggles to align central-government strategy with local implementation, but in China the problem is particularly acute. It has taken years for the government to formulate a response to the HIV epidemic that has been emerging before its eyes. Severe acute respiratory syndrome (SARS) first appeared in China in November

2002, but it took months for Beijing to learn of the problem, acknowledge it and orchestrate a response. And with regard to bird flu, officials last summer claimed to be unaware that the widespread misuse of antiviral drugs by Chinese chicken farmers had already rendered them virtually ineffective (see *Nature* 435, 1009; 2005).

With something that will move as rapidly as bird flu, China needs a surveillance system that brings accurate information to the centre. Once it has that, the information should be shared with scientists abroad. Such sharing of information is another thing to which Beijing has historically been averse. Last spring, for example, foreign scientists were allowed only limited access to sites of early outbreaks of bird flu in western China.

The preferred crisis-response mode of senior Chinese officials, ingrained over decades, is first to identify a problem, then take measures to address it, and only afterwards report the improving situation to the outside world.

"Sharing of information is something to which Beijing has historically been averse."

The outbreak of SARS was supposed to have demonstrated that this approach is no longer good enough. There are indeed signs that some lessons have been learned. Public-health provision has improved in the affected areas, and investigations into the disease have strengthened some aspects of Chinese biomedical research (see page 382).

In the case of SARS, mainland China eventually shared data on the virus responsible with scientists in Hong Kong and abroad, who were able to identify it and help to develop effective control measures. China should speed up this process for avian flu, making samples and genetic data for circulating strains immediately available. Access to outbreak areas should also be granted to international teams seeking epidemiological data. Data and samples from strains that have crossed over to humans also need to be shared. So far, this hasn't been happening. China's response to bird flu has been a mixture of secrecy and parsimony that does little to serve the interests of its farmers or its people, and is not becoming of an emerging world power. ■

A firm foundation?

After more than thirty years, a European science agency is struggling to establish a clear identity.

The European Science Foundation (ESF) was born before its time, in 1974 — and has suffered for it. It could have evolved into Europe's answer to the US National Science Foundation, but history decreed otherwise. Instead, it has pottered gently along,

doing small things reasonably well. It has established high-quality research networks and conferences, and has published, on an ad-hoc basis, thoughtful documents on such policy issues as Europe's need for synchrotron facilities. But it has never really established a strong identity, and repeated exercises to carve out a lead role in the fragmented drama of European science have tended to fizzle out.

The latest such exercise, conducted last year at the request of its member organizations, may yet meet the same fate. It produced a somewhat watery-sounding mission statement that the ESF should serve as "a common platform for its member organisations in order



to advance European research and explore new directions for research at the European level". In practice this means little change. But the foundation promises a stronger focus on producing strategic 'Forward Looks' reports, which will analyse Europe's future scientific needs. Another focus will be the management of research activities generated by other organizations, such as the European Commission's Eurobiofund (see *Nature* **439**, 244; 2006).

Any grander hopes have been extinguished over the years by the steady expansion of the European research commission, allied with the inertia of the ESF's owners — some 78 national research agencies and sundry academies of science, from 30 countries.

Despite their recognition that basic scientific research needs to be supported at the European level, ESF members failed to deliver the money that it needed to do the job. Given Europe's complex regional politics and the ever-present conflict between national and European interests, it is hard to imagine how things could have turned out differently.

Instead, the commission — whose mission has never directly included basic research — has stepped into the breach. Its vigorous and well-funded programmes for networking researchers, tangled

in red tape though they may be, dwarfed the ESF's efforts. Now the creation of the European Research Council is removing the central project that would have given the ESF its most natural *raison d'être*.

The ESF, whose executive staff of almost 100 are ensconced next to the picturesque canals of Strasbourg, may yet find its forte. If the 'Forward Looks' reports are good enough, they could become influential, like those of the US National Research Council (NRC), which has no European peer. However, the NRC's reputation sits on a firm bedrock of independence — the National Academies — and it has a reliable stream of money to support its large technical staff. The ESF lacks either asset.

Europe has yet to develop a stable system of scientific institutions whose combined authority will safeguard the general well-being of European-level basic research. The ESF's new mission statement is an attempt to identify its place in that evolving system. But its member organizations must remain alert and flexible if the ESF is to remain fit enough to survive. ■

"The European Science Foundation's members have failed to deliver the money that it needed to do the job."

Warming to economics

Climate research can only gain from closer collaboration with economists.

When a small group of economists first became interested in global warming around 1990, some climate scientists viewed their activities as an unnecessary intrusion. The warming of the planet, they reasoned, was being driven by economic expansion, and action on greenhouse-gas emissions was being opposed by some governments on economic grounds. Why should economists get involved when the future of the planet was at stake?

After some years of mutual apprehension, however, collaboration between climate scientists and economists is becoming widespread (see page 374). There is a growing acceptance on the part of the scientists that economic models are an essential piece of the toolkit needed to predict climate change. And there's an acknowledgement that mitigation actions cannot fly in the face of the considerations that economists probe, such as China's desire to grow and people's desire to drive their own cars. The centrality of economics to decisions in this sphere is neatly encapsulated by the British government's decision to give the Treasury lead responsibility for climate-change policy.

The inclusive nature of the Intergovernmental Panel on Climate Change (IPCC), which is currently preparing its fourth assessment of global warming for publication late next year, has further encouraged economists and climate scientists to work together. The resulting effort has shed useful light on how economic growth, lifestyle changes, international trade, and investment in the energy sector might influence greenhouse-gas emissions. But that's the easy part. The hard part is untangling the impacts of various possible mitigation actions in order to credibly estimate their actual impact on economics, climate and society.

Unfortunately, for the purposes of its impending fourth assessment, the IPCC won't manage to incorporate the economists' latest thinking on these different 'emissions scenarios'. The 'Special Report on Emissions Scenarios' that will accompany the assessment was developed in the late 1990s and rests on a number of assumptions that many economists view as outdated or simplistic.

For example, it assumes direct 'cause-and-effect' correlations between factors such as population growth and technological change, instead of the more complex, two-way relationships that economists have established beyond reasonable doubt. It also makes macroeconomic assumptions, such as a rapid convergence between the per-capita income of rich and poor nations, that ought really to be discarded as wishful thinking.

Common-sense adjustments to the report on emissions scenarios could incorporate the best understanding that we have regarding the interaction of such variables in complex economic systems. They will not remove uncertainty, which is fiercely ingrained in economics. Increasingly empirical approaches to economic assessment and modelling, many of them borrowed from the 'hard' sciences, will, one suspects, never lead to a mechanistic understanding of economic forces. The development of such approaches is welcome nonetheless, and has an important role to play in the assessment of emissions scenarios.

The IPCC has initiated the development of an improved set of such scenarios for completion by the end of the decade, in time for incorporation into its fifth climate assessment, due in 2013. Such are the slow wheels of progress at an organization designed to forge painstaking consensus. The delay need not undermine the authority of the IPCC's work, but it will doubtless lend ammunition to its vocal and well-financed critics when the fourth assessment is released. ■

"Unfortunately, for the purposes of its impending fourth assessment, the IPCC won't manage to incorporate economists' latest thinking."

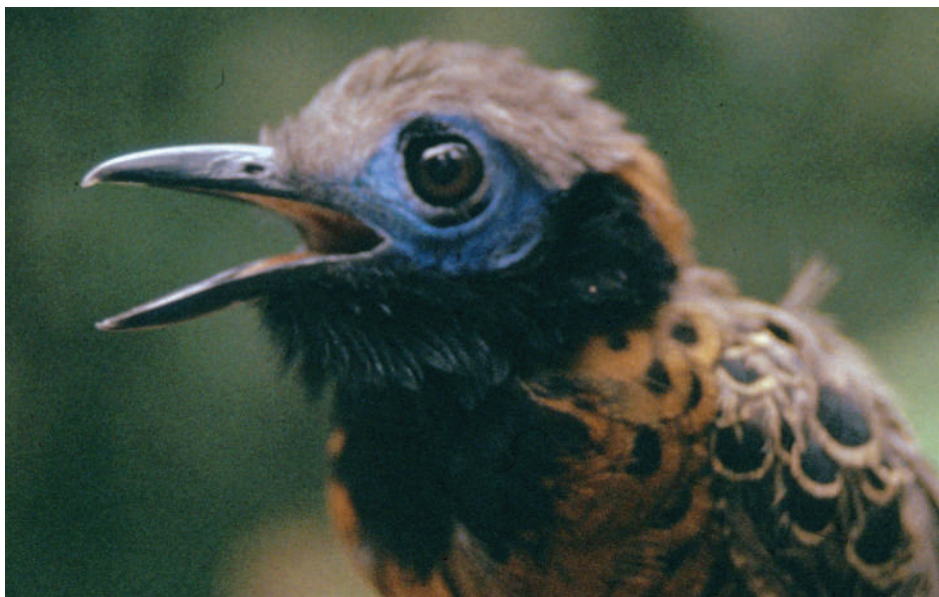
RESEARCH HIGHLIGHTS

Singing in the rainforest

Proc. R. Soc. Lond. B doi:10.1098/rspb.2005.3410 (2006)

What determines when each bird species joins the dawn chorus? To find out, Karl Berg of Florida International University in Miami and his colleagues spent a month recording songbirds in the tropical forests of lowland Ecuador, home to species such as the ocellated antbird (*Phaenostictus mcleannani*, pictured).

The researchers collected data on 57 species. The songbirds join the chorus in a chronological sequence determined by the height at which they roost in the 20–25-metre-tall forest, and by the size of their eyes. This fits with the hypothesis that birds start to sing when conditions are bright enough for them to spot predators drawn by the noise, continuing until the light allows them to find food themselves.



K. S. BERG

CONDENSED MATTER

Highs and lows

Phys. Rev. Lett. **96**, 017006 (2006)

For decades, physicists have been trying to convert pure hydrogen into a metal by compressing it. Theory predicts that this feat should be possible. However, despite reaching pressures exceeding 300 gigapascals — about 3 million atmospheres — they have failed.

Now Roald Hoffmann of Cornell University in New York and five other theorists have explored an alternative approach, which involves compressing molecules that contain hydrogen, such as silane (SiH_4).

They analysed 13 possible crystal structures for solid SiH_4 , concluding that it should become a superconducting metal at experimentally accessible pressures, possibly as low as 91 gigapascals.

CELL BIOLOGY

Into your arms

J. Cell Sci. **119**, 370–379 (2006)

A moving amoeba may help researchers to uncover the function of a gene that is mutated in a rare human disease.

Patients with Shwachman–Bodian–Diamond syndrome (SBDS) can suffer problems with their immune system. It has also been observed that patients' white blood cells do not manoeuvre normally in response to chemical signals.

David Soll of the University of Iowa in Iowa City and his team identify the homologue of the SBDS gene in the amoeba *Dictyostelium discoideum*, and show that the protein it

makes travels into the organism's protruding arms when it migrates in a chemical gradient.

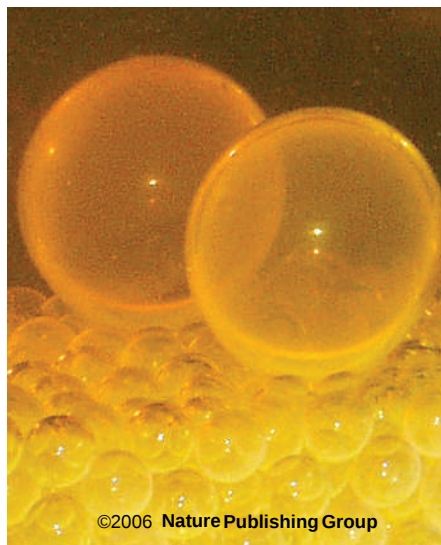
The protein may carry RNA molecules required for cell movement, they speculate, and failure of this process might underlie some symptoms of the human disease.

BACTERIOLOGY

Magnetic reversal

Science **311**, 371–374 (2006)

Some bacteria have internal compasses, made from a chain of magnetic iron crystals. Until now, everyone thought that Northern-Hemisphere bacteria used their compasses to swim north, and Southern-Hemisphere bacteria used theirs to swim south. In either half of the globe, this would allow the bacteria to squirm into the lower-oxygen depths of water bodies by following the downward slope of the Earth's magnetic field.



But Katrina Edwards of the Woods Hole Oceanographic Institution and two colleagues have smashed this simple explanation with the discovery of dumbbell-shaped bacteria that flout convention. The iconoclasts, collected from a saltwater basin in the Northern Hemisphere, swim resolutely south under the microscope. The next thing to do is find out why.

FLUID DYNAMICS

Not a drop

Langmuir **22**, 522–523 (2006)

Behold the antidrop! The two large bubbles in the picture (left) constitute a thin shell of salty water, sitting in and enclosing a mixture of oil and emulsifier. They are about eight millimetres across and rest on a bed of ordinary water droplets.

Discovered accidentally by a team of scientists in the United Kingdom and Australia, the antidrops are not to be confused with 'antibubbles', which are spherical films of air in water. The team, led by Kevin Galvin of the University of Newcastle, New South Wales, produced the antidrops by introducing oil through a perforated plate at the bottom of a vessel — the oil acquires an aqueous shell as it rises through the water drops.

RNA INTERFERENCE

Silent advance

Genes Dev. **20**, 147–152 (2006)

'Inhibitory RNAs', a recently discovered class of RNA, have huge potential as tools in biology and medicine because they can

inhibit the expression of specific genes.

Miles Wilkinson from the MD Anderson Cancer Center in Houston, Texas, and his colleagues have developed a method of RNA interference analysis that allows a specific gene to be suppressed in just one tissue in transgenic mice. Their trick exploits the way that naturally occurring gene-suppressing microRNAs are produced.

They applied the technique to demonstrate that Wilm's Tumour 1 transcription factor, known to be required for normal development of mouse embryos, also plays an essential role in sperm production and motility in adult mice.

MALARIA

Once bitten

Nature Med. doi:10.1038/nm1350 (2006)

Mosquito bites introduce the malaria-causing *Plasmodium* parasite deep into the skin. A study in mice has mapped its subsequent path.

Robert Ménard of the Pasteur Institute in Paris, France, and his colleagues tracked the fate of the rodent-infecting *Plasmodium berghei*, modified to express green fluorescent protein.

Their real-time analysis reveals that, within an hour of a bite by an infected mosquito, roughly 50% of the cells injected into the mouse, known as sporozoites, have left the skin. Of these, about 70% enter blood capillaries and thus can lead to disease, whereas 30% go into lymphatic vessels.

The next two findings both came as a surprise. Most of the sporozoites entering the lymph system were degraded in lymph nodes. Also, some of the sporozoites in the lymph matured to a stage at which they could influence the immune response to the infection.



MATERIALS SCIENCE

The complete Plato

Small doi:10.1002/sml.200500362 (2006)

According to Plato, the heavenly ether and the classical elements — earth, air, fire and water — were composed of atoms shaped like polyhedra whose faces are identical, regular polygons. Such shapes are now known as the Platonic solids, of which there are five: the tetrahedron, cube, octahedron, icosahedron and dodecahedron. Microscopic clusters of atoms have already been identified with all of these shapes except the last.

Now, researchers led by José Luis Rodríguez-López of the Institute for Scientific and Technological Research of San Luis Potosí in Mexico and Miguel José-Yacamán of the University of Texas, Austin, complete the set. They find that clusters of a gold–palladium alloy about two nanometres across can adopt a dodecahedral shape.

GENETICS

Two trees, side by side

Genetics doi:10.1534/genetics.105.048439 (2006)

Comparing the genomes of oak and chestnut trees outlines the story of their evolution since the two species split 60 million years ago.

Antoine Kremer of the French Institute for Agricultural Research near Bordeaux and his colleagues have sequenced and compared DNA from oak (*Quercus robur*) and chestnut (*Castanea sativa*, pictured left). They looked at sites called quantitative trait loci (QTL), which are associated with heritable traits.

Of the 13 sites known to influence the timing of bud burst in oak, and 10 in chestnut, 9 are found to match. But for other traits, including one that helps to govern drought tolerance, the two tree species do not have any QTL in common, showing where they have faced different selection pressures.

CELL BIOLOGY

Centre of attention

Science **311**, 388–391 (2006)

Chromosomes play an active role in positioning themselves centrally during cell division, say researchers.

The replicated DNA molecules within each chromosome are pulled apart during cell division by two sets of molecular fibres that extend from opposite ends of the cell. This happens in two stages, with the fibres from one side attaching first, and dragging the chromosomes to their side of the cell. It was unclear how the second set of fibres could then locate the chromosomes on the far side of the cell. But it turns out they don't have to.

Alexey Khodjakov of the Wadsworth Center in Albany, New York, and his colleagues show that mammalian chromosomes make themselves easier to find. With the help of a motor protein known as CENP-E (kinesin-7), they travel along the fibres of already centred chromosomes — rather like using a handcar to travel along a railroad — to reach the middle of the cell.

JOURNAL CLUB

Mary Schweitzer
North Carolina State
University, Raleigh

As new fossil evidence is uncovered, old prejudices fly out of the window, says a palaeontologist.

When the debate over the value of molecular versus morphological data in elucidating evolutionary relationships was in its infancy, one of my colleagues published molecular evidence suggesting that

modern birds had emerged during the era of the dinosaurs — much earlier than fossil evidence suggested (B. Hedges *et al.* *Nature* **381**, 226–229; 1996).

Shortly thereafter, I attended a meeting of palaeontologists at which the paper and its implications were disputed, loudly. I overheard one delegate suggest, with annoyance, that molecular people should stick to living organisms, and leave the discerning of early evolutionary patterns to those who study fossils.

Molecular data, from the

palaeontologist's point of view, is fine if it agrees with the fossil evidence. But there has been resistance when it contradicts conventional views.

As one who dabbles in palaeontology and molecular biology, I was happy to see a small fossil reconcile the approaches in the debate over birds.

Julia Clarke of North Carolina State University and her colleagues recently analysed the fossil remains of an Antarctic bird dating to the Late Cretaceous period, around 70 million years ago (J. Clarke *et al.*

Nature **433**, 305–308; 2005). They identify the bird as a new species, *Vegavis iaai*, which belongs to a lineage that includes today's ducks.

The bird shows specialized traits, suggesting that its lineage diverged long before this specimen lived, and well before the extinction of the dinosaurs, consistent with Hedges' hypothesis. The fossil provides independent evidence to support the molecular data, and acts as a stunning reminder that we must be prepared to re-examine our conventional wisdom in the light of new data.

SPECIAL REPORT

The costs of global warming

Efforts to forecast how Earth's future climate will affect us must consider the economic growth of both rich and poor nations. But there are doubts over the theories being used, as **Quirin Schiermeier** explains.

Discussions of climate change tend to involve uncertainties, and most climate researchers have come to accept the inherent unknowns of their business. After all, the climate models they use to project the course of global warming are generally seen as the best that science can offer. But there is a growing feeling that the economic assumptions on which their work is based are outdated and unreliable. And this could have serious implications for assessments of climate change.

The Intergovernmental Panel on Climate Change (IPCC), which coordinates efforts to predict the effects of global warming, is currently finalizing its fourth assessment report. It has asked 15 climate groups to run their models using output from a range of different 'scenarios', representing various assumptions about energy use, economic development and population increase up to 2100.

For the different models to be compared, all groups will run two of the 40 available scenarios — A1B, which describes a world with large economic growth and a balance between fossil and renewable energy sources, and B1, which assumes less economic growth and a large shift towards renewable energy. Each group will also run a selection of other scenarios. The results will be fed into the report of the IPCC's Working Group I, which projects how temperatures may rise over the next century.

Working Group II will translate this into likely effects on society, income, food security and disease, and Working Group III will assess the options for limiting greenhouse-gas emissions and mitigating climate change.

The scenarios were described in the IPCC's "Special Report on Emissions Scenarios" (SRES) for the panel's 2001 third assessment report. In June 2003, the IPCC decided that there was no time to develop new scenarios for its fourth assessment report, so the same ones are being used again.

But many economists are vehemently questioning the assumptions on which the SRES scenarios are based. They say the scenarios rely on outdated economic theories and fail to

reflect how lifestyle and energy demand in both rich and poor countries are likely to change.

Climate researchers are familiar with the problem. "Some emissions scenarios are perhaps already demonstrably wrong," says Erich Roeckner, a climate modeller at the Max Planck Institute for Meteorology in Hamburg, Germany, who has modelled three of them for the IPCC (see "Early results"). "It is possible that all of them are wrong." But most feel that economics is a field they are not qualified to assess.

Ridiculous assumption?

One key criticism is the assumption that the economies of poor countries will quickly catch up with those of rich nations. "It is ridiculous to assume, as the IPCC does, that rich and poor countries will economically converge as rapidly as the European Union has done over the past 40 years," says Richard Tol, an economist at Princeton University in New Jersey.

A particular worry about convergence was first raised by David Henderson, former head of economics at the Organisation for Economic Co-operation and Development, and Ian Castles, former head of Australia's bureau of statistics, after the IPCC's third assessment report. They pointed out that the IPCC used market exchange rates to compare the wealth of different countries, instead of what a certain amount of money will buy in each place. This exaggerates the differences between countries, they argue, and overestimates how much poor countries will develop in coming decades, and how much their carbon emissions will increase.

Tol says that even if certain economic assumptions are wrong, the predictions of global warming itself may not be far off. The errors may even cancel each other out: if economic development is slower than thought, countries may also switch to renewable energy more slowly.

Susan Solomon, who chairs the IPCC's Working Group I, emphasizes this point. "Existing scenarios cover a very wide range of emissions trajectories," she says. "They are perfectly suited to physical tests of how the climate



Translating temperature changes into impacts on society is beset with unknowns.

responds to fixed concentrations of greenhouse gas. In the simplified view of science it does not matter how the gas gets into the atmosphere, only how much of it is there."

But Tol argues that when it comes to translating those temperature predictions into impacts on society, how the carbon got there matters very much indeed. Many problems relating to climate change, such as the distribution of malaria and water-borne diseases, are highly sensitive to development and wealth, says Tol. "You can't use flawed economic scenarios for any meaningful analysis of the impacts of climate change."

In terms of human welfare, some even assert that the difference between the economic effects of the various scenarios is more significant than the predicted effects of climate change. For example, changes in malaria incidence that result from people becoming rich enough to have mosquito nets outweigh any changes in the geographical spread of malaria caused by global warming.

Development of the SRES scenarios was overseen by Nebojsa Nakicenovic, an energy economist at Vienna Technical University in Austria. He says he is aware of the criticisms, but argues that the scenarios were designed to



HALT CALLED ON SINGLE-DRUG ANTIMALARIAL PRESCRIPTIONS

Irresponsible treatments could render artemisinin ineffective.

www.nature.com/news



cover a range of possibilities, and include one group that assumes little convergence.

Martin Parry, co-chair of Working Group II, which met in Merida, Mexico, last week to consider an early draft of the fourth assessment report, adds that he and his colleagues also take other projections into account, for example those from the World Bank or the Millennium Ecosystem Assessment. But which ones are likely to be right isn't for them to judge, says Parry. "The IPCC summarizes current knowledge — it does not have a view."

But economists such as Tol say that instead

of considering an ever wider range of possibilities, the IPCC needs to rethink its scenarios and come up with a selection of likely options based on more up-to-date economic theory.

Tol believes that the composition of the SRES teams was too narrow, and that future efforts should have a wide range of experts. As well as convergence, they will need to consider how future societies will operate, how fast the population will grow, and how technological progress will change things. "All these questions are at the heart of economics," agrees Ottmar Edenhofer, an economist at the Potsdam Institute for Climate Impact Research in Germany.

Edenhofer says that economists must also routinely check their ideas against historical data, and test their predictions against those of other teams, as climate scientists do. The way natural scientists compare their models "is a totally new thing in economics", says Edenhofer. "We can learn a lot from climate modellers."

The aim is to develop flexible theories, in which variables such as gross domestic product and technical progress influence one another, to test the effects of policy interventions — emissions caps or trading, for example — and work out the cost of different policies.

A few attempts have already been completed. In 2003, economists Michael Grubb of Imperial College London and Carlo Carraro of the University of Venice, along with John Schellnhuber, research director at the Tyndall Centre for Climate Change Research in Norwich, UK, launched a project to investigate how technological progress affects emissions strategies and mitigation costs.

Their comparison of models incorporating data on the economy, energy use and the environment suggests that climate policies such as the Kyoto Protocol will trigger technological

change. They estimate that the cost of stabilizing emissions at 450 parts per million — a value it is hoped will be sufficient to avoid dangerous climate change — would be just 0.4% of the world's present gross product (O. Edenhofer *et al.* *Energy J.*, in the press).

And this week, William Nordhaus, an economist at Yale University in New Haven, Connecticut, published results from a model that combines data on historical climate, economic activity and country-specific temperature predictions to estimate the economic impact of a temperature rise of 3 °C. He concludes that such a rise would cause a 1–3% decline in global income — significantly larger than previous estimates (W. D. Nordhaus

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0509842103; 2006).

The IPCC is starting to consider economists' concerns. In September 2005 it set up a group to look at scenario development led by Bert Metz, a chemist at the Netherlands Environmental Assessment Agency and co-chair of Working Group III. Metz says he will recommend the development of completely new emissions scenarios, including a more sophisticated treatment of convergence and technology. The IPCC will decide on whether to go ahead with Metz's plan at a meeting in April on Mauritius. But even if work starts now, the scenarios won't be ready for use until the IPCC's fifth assessment report, due in 2013 or later.

For critics such as Tol and Edenhofer, that's not a moment too soon. "Economists must learn to bring their knowledge to bear in the discussion, and climate scientists need to realize the significance of economics," says Edenhofer. Rather than arguing about temperature predictions, "the future battle will be about cost and damages".

See Editorial on page 370.

Early results

Global temperatures are likely to rise by 2.5–4 °C by 2100, according to the latest calculations by scientists at the Max Planck Institute for Meteorology in Hamburg, Germany.

The institute is one of 15 asked by the Intergovernmental Panel on Climate Change to run extended climate simulations for its fourth assessment report. The researchers ran six parallel

experiments, requiring 400,000 computing hours, using their atmospheric general circulation model ECHAM5.

They looked at three emissions scenarios, representing carbon dioxide concentrations of 550, 700 and 800 parts per million (p.p.m.) by 2100 (see graph). Even under the most optimistic assumptions, the model suggests that the Arctic

will become ice-free during summer by 2090, says Erich Roeckner, who heads the group. The global sea level will rise by up to 30 centimetres as water warms and expands, and by an additional 10 centimetres as part of Greenland's ice sheet melts. The scientists also expect a weakening — but not a shut-down — of the Atlantic ocean circulation. There will be more rain and snow at high latitudes and in

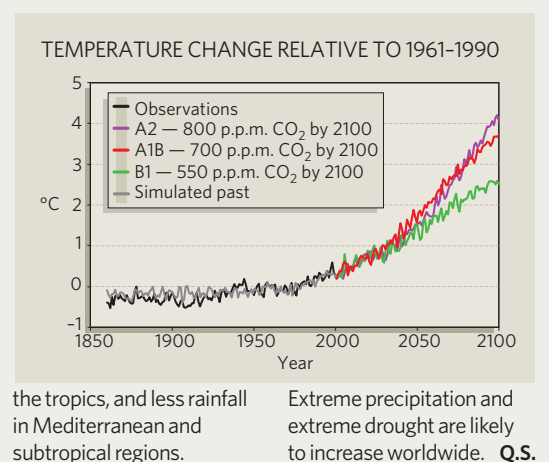


IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The deaths of beaked whales
have been blamed on naval
exercises involving sonar.

Panel quits in row over sonar damage

SAN DIEGO

As the US Navy plans to expand its testing of sonar in the oceans, the creation of an independent research programme to find out how the noise may affect marine mammals seems doomed. Evidence is also emerging that the Navy may have been pressuring scientists to downplay links between sonar and damage to marine life.

After two years of meetings costing nearly \$1 million, an advisory committee of scientific, military and industry leaders convened by the US Marine Mammal Commission (MMC) has collapsed. Instead of producing a consensus-based report on how best to study the effects of sonar on marine mammals, the 28 members will, next week, submit individual recommendations. The MMC will then report to US Congress in the next few months, but without agreement, it is unlikely that any research will be funded.

At the beginning of the decade, environmental groups took the Navy to court over its use of low-frequency sonar. The Navy lost early court rounds, but after going to Congress, won exemptions from environmental laws. In 2003, Congress created the MMC advisory panel, in

which warring parties were brought together to hammer out a plan for future research and management. It was hoped that agreed research questions would be pursued by an independent programme, estimated to cost about \$25 million over five years.

Such research is badly needed. Little is known about how marine mammals are affected by sonar — although whales or dolphins have repeatedly been found beached after military sonar tests. The strongest evidence for its destructive effects comes from British researchers, who reported that military sonar off the Canary Islands was linked to decompression deaths of beaked whales (P. D. Jepson *et al. Nature* **425**, 575–576; 2003). A subsequent study found cavities in sperm whale skeletons, supporting the idea that whales suffer from decompression sickness (M. J. Moore and G. A. Early *Science* **306**, 2215; 2004).

MMC executive director David Cottingham says the committee couldn't agree because of "the high degree of uncertainty over the impact of various noises on marine mammals". But interviews with observers and panel members suggest that the Navy, as well as other groups that use sonar, including geophysical

researchers and the oil and gas industry, blocked a consensus. A Navy spokesman, however, denies this; along with representatives from the other groups, Navy officials insist that they are interested in good science.

What lies beneath

Mammalogists on the panel disagree. "This process has been a travesty of fiscal responsibility, scientific integrity, and environmental stewardship," Lindy Weilgart of Dalhousie University in Nova Scotia, Canada, wrote to the MMC as the committee disintegrated last September.

"The science of ocean sound is highly politicized," adds Hal Whitehead, a marine mammalogist at Dalhousie and Weilgart's husband. "I see the breakdown of the committee as an indication that the Navy and others didn't want Congress to have a clear picture of what the risks are."

Either way, the promised research is unlikely to happen. And scientists question whether the current US programme, funded mainly by the Navy, will tackle questions fairly and openly. Late last year, a lawsuit forced the release of e-mails in which military officials discussed their attempts to pressurize a

researcher funded by the Office of Naval Research (ONR) to withhold comments on the damaging effects of sonar. The 2001 e-mails detail how Robert Gisiner, who manages the ONR's mammal research funding programme, engaged in "a pretty scorching phone call" with Robert Schusterman, a marine biologist at the University of California at Santa Cruz. Schusterman had filed comments for an environmental report saying that a Navy sonar test could be harmful to marine mammals. Gisiner denies any impropriety and says he was simply "talking to an old friend".

And last week, a report by Teri Rowles, coordinator of the National Marine Fisheries Service's (NMFS) stranding programme, was made public in another court case; the National Resources Defense Council is aiming to force the NMFS to release information about the

potential impact of a new training range planned off North Carolina. In an initial version of the report, Rowles reported that the death of at least one whale stranded on the North Carolina coast last year could have been caused by sonar. But in the final report released by the NMFS, the link to sonar had been removed.

Critics see such incidents as evidence of the conflict of interest inherent in the current Navy programme. Whitehead and Weilgart wrote in October that the funding system should be changed, "to safeguard the credibility of the field and protect us all from conflicts of interest" (see *Mar. Mamm. Sci.* 21, 779–781; 2005).

Meanwhile, the public has until the end of January to comment on the Navy's plans for a new training range. ■

Rex Dalton

Nations wrestle to host future telescope

Observatories in Chile and Mexico are vying to host a new kind of astronomical facility, which, if approved, will set the standard for sensitive all-sky surveys in the next decade.

The Large Synoptic Survey Telescope (LSST) corporation, based in Tucson, Arizona, has asked for proposals from astronomers at three sites chosen for their excellent visibility: Las Campanas and Cerro Pachón in Chile, and San Pedro Mártir in Mexico. The winning site will be announced in April.

The LSST's ambitious goal is a kind of "celestial cinematograph", says project director Tony Tyson of the University of California, Davis. A single 8.4-metre telescope will photograph the entire sky every three days, across a range of wavelengths from ultraviolet to near-infrared.

The survey will pick up far fainter objects than today's Sloan Digital Sky Survey, and LSST images could be used for a range of astronomical problems, from searching for dark matter to tracking fast-moving asteroids. The volume of data generated will be unprecedented — 20 terabytes

of raw image data, more than the entire Sloan Survey's output, every night.

Tyson sees the LSST as a "totally different paradigm" for astronomy, akin to the Human Genome Project. All the data will be shared freely, as will the algorithms for image



Mexico's San Pedro Mártir site is favoured for its remoteness.

processing. He says the project is now spending most of its time on software development, and recently used Wayne Rosing, Google's former senior vice-president of engineering, as a consultant.

Private funding has allowed work to begin on the mirrors, and the National Science Foundation (NSF) is providing US\$14.2 million over four years. Still, that's far from the \$270 million

needed for construction and a decade of operations. Tyson hopes the US Department of Energy and the NSF will contribute roughly \$100 million each, with the rest coming from private sources.

The Chilean sites already have several premier telescopes, including the Gemini and SOAR (Southern Astrophysical Research Telescope) on Cerro Pachón and the Magellan telescopes on Las Campanas. The San Pedro Mártir site in the Mexican state of Baja California is less well known; its largest instrument is a modest 2.1 metres in diameter.

But the remoteness of the Mexican site could be a bonus; the Chilean observatories have had to contend with light pollution as the population builds up in nearby communities. The 'astro-climate' for the three sites — which includes factors from atmospheric transparency and stability to the number of clear nights per year — is virtually identical, says Tyson. So other considerations, from local infrastructure to politics, may sway the decision. ■
Tony Reichhardt

Treasure island: pinning down a model ecosystem

Lounging on Craig Venter's yacht in the South Pacific a couple of years ago, Neil Davies contemplated the tiny island of Moorea. Venter, famous for his work on the human genome, was sailing around the world to catalogue the microscopic life of the oceans. But Davies was pondering a more audacious goal: a plan to sequence an entire island. He mentioned the plan to a scientist on Venter's crew: "He just laughed," Davies remembers.

But Davies was on to something. He and a band of ecologists are launching the Moorea Biocode Project, which aims to turn the island into something like a model organism for tropical ecology. Christopher Meyer, a sea-snail expert at the University of Florida, Gainesville, is the plan's coordinator. Meyer says it will build on the ideas and technologies behind the scientific movement known as DNA barcoding, which classifies species according to a specific stretch of their genetic sequence. But Meyer says the Moorea project will go further.

He and his colleagues plan to collect multiple genetic and ecological data about each species on Moorea, which lies 15 km northwest of Tahiti. They will deposit the information in linked databases. Meyer hopes this will give scientists more information than barcoding a single DNA sequence. "Barcoding is great, and a lot of people are excited about it, but it can only answer questions about one narrow space," Meyer says. "We intend to fill our data set with additional information so that we can answer a broader set of questions."

"Barcoding is great, and a lot of people are excited about it, but we intend to answer broader questions."

The Biocode team will meet next month to begin designing databases. Meanwhile, entomologist Rosie Gillespie — based, like Davies, at the University of California, Berkeley — will begin collecting insects on Moorea and sequencing their DNA. Then, in March, French ichthyologist Serge Planes of Perpignan University will start sampling fish from Moorea's reefs; he hopes to collect 80% of the 600 fish species in four-and-a-half weeks. Researchers intend to start using these data immediately to look at topics from invasive species to biodiversity.

Nancy Knowlton, a coral-reef expert not involved with the Biocode project, says such data could resolve many unanswered questions. For instance, she says, it is often hard to identify tropical reef fishes, many of which have been described only in small journals. Having a DNA code linked to a visual key could

help biologists to make quicker, more accurate identifications. And that could help them to understand crucial parts of reef ecosystems, such as how many species live on them and how well they are doing.

"Our estimate of the number of species on reefs rests on incredibly shaky ground," says Knowlton, who directs the Center for Marine Biodiversity and Conservation at Scripps Institution of Oceanography in La Jolla, California. "These molecular tools have the potential to help us fine-tune those estimates to get a total sense of diversity, and what we're losing as the reefs degrade."

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Pacific crews: collecting genetic data from all life on Moorea should shed light on biodiversity.

The Biocode scientists also want to learn about the general properties of ecosystems. Moorea is less diverse than other islands farther east, so it may serve as a reference site that can be compared with more complex systems in Australia, Papua New Guinea and southeast Asia. "It's like comparing the processes of *Caenorhabditis elegans* with humans; that's a very powerful approach," says Davies.

Moorea is a logical choice for a model system of ecology, experts say, because it has been

1930

Clyde Tombaugh discovers Pluto while searching for 'planet X', predicted by Percival Lowell in the late nineteenth century to explain an apparent anomaly in the orbit of Uranus.

1976

Methane ice discovered on Pluto. Nitrogen and carbon monoxide ices are also identified 16 years later.

1978

Discovering Pluto's moon Charon allows astronomers to determine their masses. They find that Pluto is much smaller than they thought, and conclude that it is made of ice and rock.

1985-88

Pluto found to have a tenuous atmosphere, probably of nitrogen gas. Observations in 2003 show that the atmospheric pressure has since doubled, presumably caused by seasonal warming.

1989

Pluto makes its closest approach to the Sun, bringing it 29.7 astronomical units away (1 AU = the Earth-Sun separation). Its highly eccentric, 248-year orbit means that during 1979-99 it was closer to the Sun than Neptune. It will swing back out to its most distant point of 49.5 AU in about 2113.

1992

The first evidence that Pluto is just one of a band of numerous Kuiper-belt objects (KBOs) arrives with the discovery of (15760) 1992 QB₁. The belt had been hypothesized for decades, suggested as a repository of debris from the Solar System's formation.

A BRIEF HISTORY OF PLUTO ►

As NASA's New Horizons spacecraft sets off on its long trip to Pluto, Mark Peplow looks at how our view of the Solar System's outskirts has changed.



SPACE NEEDS AN URGENT SPRING CLEAN
Rising levels of debris are likely to cause catastrophic crashes.
www.nature.com/news

M. WADE/WWW.ASTRONAUTIX.COM



Y. ARTHUS-BERTRAND/CORBIS

well studied by researchers at two field stations there for decades. The project can also tap into other, similar efforts that are already under way, such as the Census of Marine Life, in which Knowlton is involved.

Putting a whole island under the microscope won't be easy. But the Biocode scientists say they are building on a tide of change that is revolutionizing taxonomy and ecology. "We have technological challenges; we have sampling challenges," Meyer admits. "But the idea of bar-coding has really hit a tipping point. This is the perfect time to try something like this."

Erika Check

Bad data fail to halt patents

Two patent applications filed by the California Institute of Technology will proceed despite concerns over the accuracy of data they contain, *Nature* has learned.

The patents were filed on behalf of the institute's president, David Baltimore, and Luk Van Parijs, formerly a professor at the Massachusetts Institute of Technology. Van Parijs was considered a rising star in the field of immunology, but was sacked last year for fabricating data in at least one published paper. Several of his papers use duplicate images, but none of his co-authors has been implicated in any misconduct.

Unlike scientific papers, patent applications do not depend on data but on claims to have conceived a new invention. Inventors often include data, however, to demonstrate the novelty or usefulness of their claims, or to increase the enforceability of the patent once granted.

In the United States, in contrast to many other countries, inventors must sign a declaration affirming that everything in their application is true to the best of their knowledge. The inclusion of false data, even by mistake, could be an infringement of

the oath, and thus against the law. Or it could form the basis for questioning the patent later, says Alan Grimaldi, co-chair of the intellectual-property group at Howrey law firm in Washington DC.

The applications were based on work carried out during Van Parijs's postdoc time in Baltimore's lab; the only other person on them is Lili Yang, a research scientist in the lab. They describe a novel method of gene therapy in which bone-marrow stem cells are infected with a virus to introduce extra genes, then transplanted into a patient where they produce modified immune cells. Depending on the genes inserted, the technique could treat HIV or cancer, the inventors claim.

The research upon which this is based has not been questioned. But the patent applications contain at least two instances in which identical images are labelled as coming from different cell populations; three others may also be wrongly identified. When questioned by *Nature*, Baltimore admitted one of the errors and said he would correct it; he is considering the others. He insists, however, that Van Parijs is not responsible. "The patents are fine, and we're proceeding

with them," he says.

"Somebody sent in the wrong file. It probably happened in the patent office after we sent them the data."

The status of false data in patents has already come under scrutiny this month. An application on human embryonic stem cells filed by South Korean cloner Woo Suk Hwang of Seoul National University includes data on a cell line discredited by an investigation into his work. His university has said the application will proceed after amendments have been made. If granted, however, some fear it could block patents by other researchers, or that Hwang might profit financially from others' work.

The Van Parijs applications may also raise concerns if not corrected. The gene therapy described is a promising area of research, and other work suggests the principle may well be valid. For example, Derek Sant'Angelo at the Memorial Sloan-Kettering Cancer Center in New York has made modified immune cells using a similar approach. In general, the existence of invalid data in patents — intentional or not — would be "a concern" to other inventors, says Sant'Angelo. ■
Eugenie Samuel Reich

2001-04

The first binary system, other than Pluto-Charon, is found in the Kuiper belt. When the relatively large KBOs Quaoar and Sedna follow in 2002 and 2004, some astronomers argue that Pluto is just one of the crowd, and not deserving of the title 'planet'.

July 2005

2003 UB₃₁₃ is spied by ace planet hunter Mike Brown's team and dubbed 'the tenth planet', as it seems to be even larger than Pluto. Its small moon is unveiled to the world two months later.

October 2005

Two tiny moons are found around Pluto, each between 50 and 160 kilometres across. The find supports the theory that Pluto and its satellites formed in a massive collision, rather than a capture event.

January 2006

Astronomers in Hawaii find that Pluto's surface temperature is -230°C , ten degrees cooler than Charon. The difference is blamed on the evaporation of nitrogen ice from Pluto's surface, keeping the planet cool.

19 January 2006

The New Horizons craft launches, off to probe the Kuiper belt. It will try to discover if Pluto has any geological activity, or even an internal liquid ocean. Sampling the atmosphere should help explain why it rapidly leaks into space. With the craft goes the man who started it all: a small vial contains Clyde Tombaugh's ashes.

July 2015

After a gravitational boost from Jupiter in 2007, New Horizons will have just six months where its views of Pluto and co. are better than those of the Hubble Space Telescope, and most observations will occur during flybys over a 24-hour period. A KBO encounter is planned for two or three years later.

R. HURT (PAC)

ON THE RECORD

“We at Clonaid believe that Dr Hwang has cloned human embryos and has the knowledge to develop stem-cell lines.”

Clonaid, the cloning company run by a religious cult, offers disgraced stem-cell researcher Woo Suk Hwang a job.

“I hope this means that I inherit a castle in Ireland.”

Novelist Peter Quinn reacts to a study showing more than 2 million people are descended from Irish nobility.

Sources: Clonaid, New York Times

SCORECARD

**French air**

The levels of greenhouse-gas emissions in France were lower last year than in 1990, putting the country on track to meet its commitments under the Kyoto Protocol on climate change.

**Japanese fishing**

Giant Nomura's jellyfish (*Stomolophus nomurai*), which can weigh up to 200 kilograms, are swarming the coast of Japan and paralysing normal fishing practices.

**Plant sex**

A study led by the University of Calgary has found that competition for a dwindling number of birds and bees is hindering pollination in the world's rainforests.

NUMBER CRUNCH

Obesity can be treated by 'bariatric' surgery, such as a gastric bypass. A study that estimates the frequency of such work in the United States suggests it is on the rise.

69,273 bariatric operations were performed in 2002.

588 patients were under the age of 20.

736% is the rise in the number of such operations between 1996 and 2002.

\$2 billion was the size of the bariatric industry in 2002.

Source: M. M. Davis et al. Arch Surg. 141, 71-75 (2006).



Geologists want access to Indian waters to understand the devastating seafloor earthquake in 2004.

India's ban on foreign boats hinders tsunami research

A fresh round of research missions is set to probe the seafloor rupture that triggered the devastating tsunami of 2004. But when researchers arrive in the Indian Ocean, they will find some areas are off-limits.

As with previous studies of the region, vessels will not have permission from the Indian government to enter the country's territorial waters. It is a policy that hampers research, say scientists involved, and ultimately sets back attempts to prepare for earthquakes.

The first of this year's missions set sail from Bremen in Germany on 20 January. Organized by the Federal Institute for Geosciences and Natural Resources in Hannover, the six-week mission will run studies such as seismic surveys and sediment sampling. The data gathered will boost efforts to model the earthquake and help gauge the likelihood of further major tsunamis in the area.

Other nations, including Britain and the United States, also plan to send ships, but none will be able to survey the northernmost 900 km of the 1,300-km rupture zone. This part lies in Indian waters, and researchers say they have not even attempted to ask for permission to enter. Indian science secretary Valangiman Ramamurthi told *Nature* that the ministry of defence does not allow any foreign vessels in its territory for "reasons of national security and sovereignty".

That stance is unusual. With other countries scientists say they are almost always able to reach an agreement to send research missions.

But after decades of being refused entry to Indian waters, they are resigned to relying on the limited data produced by Indian research vessels and the occasional chance to join those voyages.

David Tappin of the British Geological Survey in Nottingham, who is sailing with the German team, says access to the southern part of the 2004 rupture zone is crucial for studying the behaviour of the two plates that caused the earthquake. They interact differently in different places, he explains. In the north, the plates slip horizontally past each other. But in the south, the Indian plate is moving underneath its neighbour.

Chris Goldfinger, a marine geologist at Oregon State University in Corvallis who is organizing a trip for later this year, says lack of access is hampering attempts to understand how frequently major earthquakes have occurred in the area. "This is readily determined by geologic sampling," says Goldfinger. Such sampling is currently only scheduled for Indonesia, he says, leaving important work to a scattering of good but limited efforts under way in India.

In the absence of progress at the diplomatic level, researchers are focusing on collaborations with Indian colleagues. Goldfinger says he is working with the National Institute of Oceanography in Goa and hopes to participate in future Indian missions.

Jim Giles

Additional reporting by K. S. Jayaraman in New Delhi

Ecoterrorists arrested ahead of arson attack

Two groups of radical US environmentalists — both linked to threats or attacks on researchers — have been targeted in a crackdown on ecoterrorism.

On 13 January, three people were arrested in the parking lot of a Kmart store in California after buying supplies that could have been used to make bombs. Federal officials say the group was planning to set fire to the US Forest Service's Institute of Forest Genetics in Placerville. The lab houses 25 researchers in a facility that has been used for forest research since the 1920s.

A week later, 11 people were indicted for a series of arson attacks between 1996 and 2001. They included a 2001 bombing at a University of Washington horticulture lab in Seattle that cost millions of dollars in damages and set several research programmes back years.

Both groups have been linked to the nebulous Earth Liberation Front, an ecoterrorism group founded in Britain in the 1990s that uses sabotage to stop what it sees as exploitation of the natural environment.

Sandia takes lead role on nuclear-waste site

Sandia National Laboratories in New Mexico is to lead the research and technical work at the troubled US\$58-billion Yucca Mountain nuclear-waste repository in Nevada.

The 18 January announcement names Sandia as the successor to a consortium led by defence contractor Bechtel. Sandia, which also conducts nuclear-weapons research, will receive around \$60 million in annual funding for scientific research into the repository.

In 1987, Yucca Mountain was designated as the site of a permanent national nuclear-waste repository. Last spring, a string of e-mails from project scientists called into question the legitimacy of some environmental data taken at the site (see *Nature* 434, 427; 2005).

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

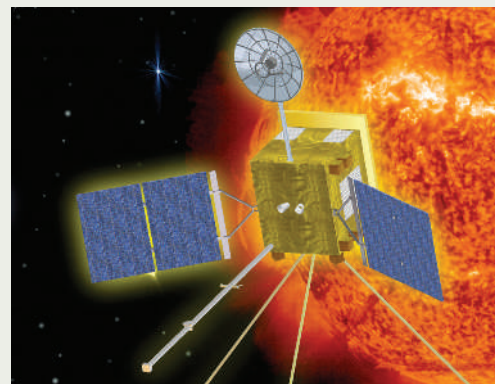
Warm welcome? Yucca Mountain science will soon be run by Sandia National Laboratories.

Europe's Sun mission eclipsed by budget shortfall

The European Space Agency should postpone its plans to send a spacecraft to the Sun, say scientists asked by the agency how it should make best use of its science budget.

A funding shortfall in the agency's science programme had led to speculation that missions may be cancelled (see *Nature* 438, 542–543; 2005). But on 19 January, its space-science advisory committee endorsed projects already in the pipeline, including the Gaia satellite and the BepiColombo mission to Mercury. It also said it will back a call for proposals for the next generation of missions.

All this is possible if the planned launch of the Solar Orbiter (pictured) is pushed back four years to 2017, says Giovanni Bignami, the committee's chairman. He will put the case to the science programme committee, which



ESA

makes the final decision, on 8–9 February.

Separately, NASA has announced that budget problems have indefinitely postponed the launch of the Dawn mission, which was to have lifted off in June to study the asteroids Ceres and Vesta.

Doctors attack constraints on access to patients' data

Medical research in Britain is being delayed or blocked by the patchwork of regulatory bodies and laws that controls access to patients' records, doctors have warned.

In a 17 January report, the Academy of Medical Sciences says that steps must be taken to help researchers deal with the array of regulatory barriers. Access to patients' records is controlled by laws including the 1998 Data Protection Act and bodies such as the General Medical Council. Studies on some issues, such as the incidence of cancer among former soldiers, have already been affected by overzealous regulation, epidemiologists warn.

The academy says regulatory bodies are being overprotective and it wants the UK Clinical Research Collaboration, which promotes research within the country's National Health Service, to take the lead in simplifying access to records.

Tobacco giant sponsors work on DNA repair

In a move that seems to break its own promise not to fund medical research, the Philip Morris Foundation has awarded €25,000 (US\$31,000) to a chemist at the University of Munich who works on DNA repair.

Anti-smoking lobbies in Germany have campaigned against the four prizes that Philip Morris awards each year in the country, arguing that the research might be used to promote the interests of the tobacco lobby. The Philip Morris Foundation insists that it should be allowed to sponsor work that has nothing to do with smoking.

DNA repair is an acceptable theme, the foundation says, because it has many applications outside medicine. The prizewinner, Thomas Carell, points out that he works nearly exclusively on plants.

Other researchers see a direct connection. Some chemicals, such as those in tobacco smoke, can cause DNA damage, which can lead to cancer if DNA repair mechanisms are overwhelmed.

Computer network focuses on marine genomes

A high-capacity computer system for analysing genomic data from marine microbes should be available to scientists by this autumn.

The creation of the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis, or CAMERA, was announced on 17 January. The \$24.5-million endeavour will be funded by the Gordon and Betty Moore Foundation.

CAMERA is a joint project of the University of California, San Diego, and the J. Craig Venter Institute in Rockville, Maryland. The university will house the computer hardware, which will be accessible worldwide through fibre-optic networks.

The system will be used to analyse the metagenomics of marine microbes — looking at microbial genomic sequence data in the context of other species.

Correction

The Editorial "Ethics and fraud" (*Nature* 439, 117–118; 2006) wrongly stated that Thereza Imanishi-Kari worked in the laboratory of David Baltimore. Imanishi-Kari had collaborated with Baltimore, but ran her own laboratory at the Massachusetts Institute of Technology.



OPEN SEASON

SARS caught China unawares. But the ensuing struggle to characterize and contain the virus has put the country's work on infectious diseases back on target. **Apoorva Mandavilli** reports.

Like anyone who was in Beijing in the spring of 2003, Hongkui Deng remembers it vividly. The Chinese government could no longer deny that the country was in the grip of a new and potentially fatal disease: severe acute respiratory syndrome (SARS). By July, the epidemic would have spread, affecting more than 8,000 people worldwide and claiming 813 lives; but in April, the panic was already palpable.

Normally bustling, the streets of Beijing were virtually deserted. The few people who ventured out wore masks and gloves, and avoided even eye contact with others. Cinemas, schools and shops were closed. It was, as many describe it, frightening and eerie — even apocalyptic. “Everyone was scared,” Deng recalls.

Deng, a cell biologist, had returned home in 2001 after more than a decade in the United States. Now based at Peking University, he was pursuing his research on embryonic stem cells. Returning from a conference in April 2003, he learnt that the mother of one of his students had SARS. Once officials had sprayed the lab, Deng's students began asking if they could work on the disease that was paralysing the nation.

“Everybody wanted to do something,” he says.

Deng had limited experience in virology, apart from a short stint working on HIV, and his students had even less. But like many other scientists in China, the team saw research on SARS as both an opportunity and a duty, and set about mastering the basics — fast.

Feverish activity

For at least six months, Deng's lab stopped working on stem cells and focused entirely on SARS. It wasn't alone. Across the country, scientists trained in protein science, anatomy, immunology and biochemistry — almost anybody who could contribute in any way — were shelving their normal projects. “Everyone was working on SARS,” says Deng. “You just had to.”

That commitment has paid off. Although China still faces a great many hurdles, its government and scientific community are becoming better prepared to combat epidemics, say some US scientists. Long after global interest in SARS has waned, Chinese scientists are still publishing important work on the disease. In September 2005, for instance, one team identified bats as a natural reservoir of the SARS

virus¹. And in October, another group reported a molecule that can inhibit the replication of a wide range of SARS-like viruses².

In part, the boom in China's infectious-disease research reflects the country's growing activity in science. Government spending on the sector in 2004 was up 16% on the previous year, outpacing the country's economic growth of about 10% per year. Since the late 1990s, the number of Chinese papers published in international journals has risen dramatically³ and the number of domestic patent applications has also gone up. But SARS was in many ways a wake-up call. It served as a grim preview of the disaster — and public humiliation — that awaited China if it did not take infectious diseases seriously⁴.

Since 2003, two new infectious-disease institutes — one set up by the Chinese Academy of Sciences, the other a joint effort with the Pasteur Institute in Paris — have been set up in China. Plans for another, a collaboration with Columbia University in New York, are expected to be announced next month. Chinese scientists say that an increasing number of students are choosing to work in infectious disease, and many scientific collaborations born during the SARS epidemic are still thriving. “Before SARS, scientists routinely didn't work together,” says Deng. “China is doing much better now.”

China is a hotbed of infectious diseases: according to agencies such as the World Health Organization (WHO), hepatitis, tuberculosis, diarrhoea, encephalitis and HIV are all prevalent at epidemic rates. And with people,

AP/EUGENE HOSHIKO

poultry and animals of all kinds living in close quarters, there are plenty of opportunities for pathogens to jump between species.

SARS was an unknown entity when it struck, and was probably spread to people by infected palm civets in China's wild-animal markets. In 2005, the country also saw an unusual cluster of infections from the pig bacterium *Streptococcus suis*, which jumped from swine to infect more than 200 people. These days, all eyes are on avian influenza, and the government has already banned poultry from markets and culled millions of birds in an attempt to prevent the virus from jumping to humans.

Slow start

Given such dangers, infectious-disease experts are delighted at the changes in China. "I think SARS prepared the ground for being willing to work with emerging infectious diseases," says Robert Webster, a leading flu expert at St Jude Children's Research Hospital in Memphis, Tennessee. Looking at the way China has handled outbreaks of bird flu, Webster says, the government seems to be much more transparent and willing to share information than it was during the SARS crisis, when it denied the existence of the disease for several months. "There are still problems out there but the situation is much improved," he says.

During the initial months of the SARS epidemic, scientists on China's mainland contributed little to the understanding of the disease. Even in May 2003, weeks after virologist Albert Osterhaus of Erasmus University in Rotterdam, the Netherlands, and his colleagues had identified a novel form of coronavirus as the cause of SARS⁵, some senior Chinese scientists still refused to acknowledge that it was the culprit, backing *Chlamydia* bacteria instead. In fact, researchers at a Chinese military institute in Beijing had come to a conclusion similar to that of Osterhaus as early as March, but internal politics reportedly stifled their finding.

At a May meeting in Beijing organized by the Chinese Academy of Sciences, Ian Lipkin witnessed the disagreements firsthand. "Even at this point, there was this argument back and forth about who was right and who was wrong," recalls Lipkin, director of Columbia University's Jerome L. and Dawn Greene Infectious Disease Laboratory. Ultimately, several government officials were sacked over the handling of the crisis.

But after its initial failures — for which China was openly chastised by the international community — things changed dramatically. Lipkin says that an enormous medical facility, intended to quarantine and treat those infected, was built about 50 kilometres outside Shanghai in just a few months.

At the May meeting, Lipkin

"These are emerging and re-emerging diseases. This is research that will be very useful."

— Zihe Rao

had presented a roadmap for the government to help it combat SARS, covering drugs, vaccines and research on the basic biology of the virus. By the time he returned to China in July, "they had virtually finished everything there was on this roadmap", he says. "I can't tell you how impressed I was. It was extraordinary."

The government also realized the need for cooperating both with the international health agencies and with its own scientists. The ministries made special concessions. They created funding pools, extended grants that were about to expire and forgave deadlines so that researchers could focus on SARS.

The incentives worked, drawing some of China's most successful scientists to the field. Among them was Zihe Rao, a structural biologist who heads the Institute of Biophysics in Beijing. By October 2003, Rao's team had published the structure of a SARS protein⁶ and has since designed a molecule that inhibits the replication of SARS and several related coronaviruses². These days, more than half of Rao's lab still works on infectious diseases. "These are emerging and re-emerging diseases. This is research that will be very useful," says Rao.

Learning curve

Will SARS come back? Quite possibly. Zhihong Hu, director of the Wuhan Institute of Virology, is tracking the SARS coronavirus in palm civets, looking in particular for a strain with the two mutations required for the virus to jump to humans⁷. She says there is some evidence that the strain is already circulating in farm civets, but is waiting for confirmation before publishing her findings.

Hu says her work is progressing slowly, in part because research on SARS now faces more obstacles than it did initially, which to some degree results from a shift in government attitudes. At the height of the epidemic, the government had few restrictions on who could work with the coronavirus. But that stopped

abruptly in April 2004, after it was discovered that the virus had escaped the previous month from the Chinese National Institute of Virology in Beijing. Although the lab was among the few in China equipped to work with dangerous pathogens, the incident prompted the government to introduce strict regulations.

During the SARS epidemic, the Chinese Center for Disease Control and Prevention (CDC) freely shared samples of the SARS virus with various scientists, including Guoping Zhao, executive director of the Chinese National Human Genome Center in Shanghai. Zhao had never worked with viruses before but he and his team quickly learned enough to track the evolution of the virus through sequence analysis. In just four months, they had completed their analysis and soon after, published a paper showing the molecular evolution of the virus during the epidemic⁸. Apart from a crash course in epidemiology, Zhao says the group learned to sequence viruses from tissue samples. This expertise, he adds, could prove useful with bird flu.

But that is if China can straighten out the remaining kinks in its research structure. Not surprisingly, the Chinese CDC and the Ministry of Agriculture keep a tight rein on SARS and avian flu viruses and allow only the few labs that have biosafety level 3 facilities to work with them. But tissue samples, such those used by Zhao, are similarly controlled.

"I think the key issue in China now is all the government agencies," says Zhao. The government may have learned to share more openly with the WHO and with international experts, but it is not open enough with Chinese scientists, he says. "The government agencies should learn how to work together with research institutes," he says. "I think Dr Webster in the United States knows much more information than I do." That may be so, but the international lines of communication are not trouble-free: the WHO, for example, has complained on several occasions that it has yet to be granted access to virus samples from last year's outbreak of bird flu.

Still, Zhao and others say that they are optimistic about China's growing expertise in infectious diseases. "I think we are learning, but it's not so easy to change in one night," Zhao says. "Things have improved a lot; I hope they improve more."

Apoorva Mandavilli is senior news editor of *Nature Medicine*.



Researchers including Ian Lipkin (far right) discuss the 2003 SARS outbreak.

1. Li, W. *et al. Science* **310**, 676–679 (2005).
2. Yang, H. *et al. PLoS Biol.* **3**, e324 (2005).
3. Cyranoski, D. *Nature* **431**, 116 (2004).
4. Ho, D. *Nature* **435**, 421–422 (2005).
5. Drosten, C. *et al. N. Engl. J. Med.* **348**, 1967–1976 (2003).
6. Yang, H. *et al. Proc. Natl. Acad. Sci. USA* **100**, 13190–13195 (2003).
7. Qu, X.-X. *et al. J. Biol. Chem.* **280**, 29588–29595 (2005).
8. The Chinese SARS Molecular Epidemiology Consortium *Science* **303**, 1666–1669 (2004).



DISCOVERY IN THE DIRT

Soil microbes are notoriously hard to culture, so how can we make the ground yield its secrets?

Virginia Gewin finds that genetic sequencing — of samples not species — may be the answer.

Leonardo da Vinci once remarked that “We know more about the movement of celestial bodies than about the soil underfoot.” You could argue that his insight still holds true the best part of 500 years later. But new genomic technologies mean that the microscopic bodies that enliven soil may be about to get the attention they deserve — if not as individuals, then as communities.

Twice as much carbon is stored in Earth's soil as exists in the plants that grow from it and the animals that depend on them. It is the soil's microbes that are responsible for recycling this carbon, and other nutrients. Living in the fractal jumble of weathered rock, mineral particles and decaying organic matter are a cast of thousands, some say millions, of species. These soil organisms occupy an endless foam of tiny niches; they purify water, detoxify harmful substances and recycle waste products. They restore carbon dioxide to the air and make the atmosphere's nitrogen available to plants. Without them, continents would be deserts — home to little more than lichen, and not much of that.

Knowing which processes soil microbes are responsible for, and how, is increasingly important to everyone from farmers to climate planners. Microbiologists want to see how the organisms communicate with each other and refine the niches in which they live. Drug

developers want to know how the soil microbes poison each other with antibiotics, and other commercially minded researchers think they could discover useful industrial enzymes or additives for biofuels. But beyond a gross understanding of inputs and outputs, the specific ecological roles of microbial species or communities in the dirt remain elusive.

The main stumbling block has been that up to 99% of soil microbes cannot be grown in laboratory cultures — the traditional way to study microorganisms. But as genetic technology has improved, it has provided ways around this. First, it managed to reveal the general level of biodiversity through the sheer quantity of different sequences, then it allowed scientists to trawl for individual genes. Now technology offers the promise of extracting not just the genomes of the creatures in the soil, but — in a sense — the genome of the soil itself.

The physical, chemical and biological complexity of the soil make this sort of study a daunting prospect, even for a hugely ambi-

tious gene-sequencer such as Craig Venter. Following his work on the human genome, Venter has been using the sequencing muscle of his institutes to look at samples from seas around the world. But soil's greater complexity makes it the ultimate challenge for such studies, and arguably the biggest open frontier in biology today.

Julian Davies, a pioneer of soil genomics at the University of British Columbia in Canada, thinks that finding out what is really going on in the soil will turn our world upside down — making that below us more wonderful than that above. “Once the diversity of the microbial world is catalogued,” says Davies, “it will make astronomy look like a pitiful science.”

The full extent of this diversity is still up for debate. But its magnitude was suggested 20 years ago by Norman Pace, a microbiologist currently at the University of Colorado, Boulder. Pace adapted techniques that Carl Woese, a microbiologist at the University of Illinois at Urbana-Champaign, had used to probe the genetic relationships between all living things. Woese's work reclassified the microbial world using variations in the sequence of a highly conserved RNA molecule; Pace applied his technique to DNA taken from the environment.

Although Pace's studies excelled at revealing the diversity in soil, they were less good at suggesting what that diversity did. “We knew

“Once the diversity of the microbial world is catalogued, it will make astronomy look like a pitiful science.”

— Julian Davies

enough to know that we were ignorant about most of the organisms, but we still couldn't get our hands on them," says Jo Handelsman, a plant pathologist at the University of Wisconsin, Madison.

The problem was partly solved as researchers refined methods for taking genes from the vast number of organisms they couldn't culture and inserting them into lab workhorses such as *Escherichia coli* — allowing analysis of particular genes and functions. Handelsman dubbed these studies of DNA from uncultured organisms 'metagenomics', and was one of the first to take advantage of it.

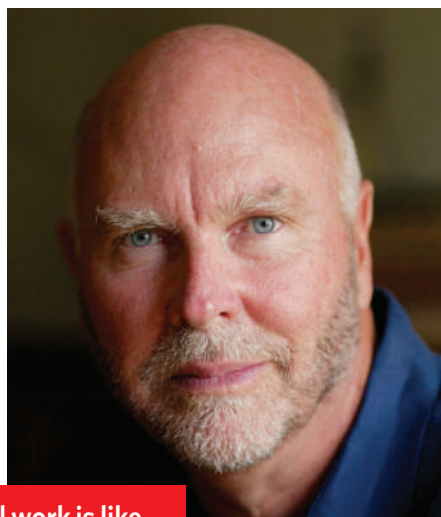
Because many antibiotics have been derived from soil microbes that do grow in cultures, such drugs were an obvious thing to look for in those that do not. So far, Handelsman and her collaborators have turned up antibiotics called turbomycin A and B (ref. 1), and Davies has discovered one he named terragine². Companies were swiftly set up to capitalize on the promise of 'functional screens' for possible antibiotics or industrial compounds; they included Diversa, based in San Diego, California, and TerraGen Discovery — now part of Cubist Pharmaceuticals in Lexington, Massachusetts.

But none of the dozen or so compounds found through metagenomic studies of soil seems set to be useful. Davies, who founded TerraGen in 1996, resignedly describes the situation as "rather disappointing". One of the problems is that the screens ignore a lot of the soil's natural diversity; genes from many of the microbes may simply not be expressed when cloned into *E. coli*. In fact, terragine is one of them: it was isolated from genes inserted into recombinant streptomycetes.

The utility of resistance

Although the search continues for new antibiotics, using metagenomics to understand soil's role in antibiotic resistance holds perhaps greater promise. Indeed, Davies has a theory that antibiotics evolved as signalling molecules that drove the development of sensing and evasion strategies in microbes. The idea is gaining popularity and leading to exciting links between environmental and medical microbiology. Recent work by Gerard Wright, a microbiologist at McMaster University in Ontario, demonstrates that soil microbes have amassed a bevy of tactics to elude the onslaught of antibiotics. On average, every one of 480 strains of *Streptomyces* that Wright harvested from soils across Canada was able to survive 7 or 8 of the 21 antibiotics presented to it — to many of which it had probably never been exposed³. Tracing genes for antibiotic resistance is an area that is ripe for study by metagenomics, enthuses Handelsman.

The search for enzymes to improve the specificity, efficiency and sustainability of



"The soil work is like trying to assemble 40,000 human genomes simultaneously."
— Craig Venter

industrial reactions has already had some success. In one desert soil sample, researchers at Diversa unearthed more than 100 enzymes that

cut up esters and lipids for use in such reactions. With only 200 esterase enzymes known before this work, discovering half as many again is no small accomplishment. For the past five years, Diversa has turned its attention to developing products, but many sit stuck in a bottleneck of government regulations, as approval processes can take several years to navigate.

Of course, the functions of genes found in soil do not have to be commercially useful in order to be ecologically enlightening. In simpler environments, such as the sea, metagenomic studies have yielded remarkable discoveries — for example, microbes that make use of sunlight without the benefit of chlorophyll. These organisms are no marginal oddity, having recently been shown to account for 13% of the microbes in the shallow parts of the ocean⁴.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Work in progress: *Streptomyces* can express the genes of soil organisms that cannot be cultured.

Soil ecologists may be envious of such a breakthrough in marine systems, but recent findings suggest that soil's secrets may not remain hidden much longer. Christa Schleper is a microbial ecologist at the University of Bergen, Norway, with a laboratory that produces libraries of large DNA fragments. She recently discovered that some of the oxidation of ammonia that goes on in the soil is being carried out by genes that come not from bacteria, but from archaea. The archaea were established as a separate kingdom of life by Woese, and until recently they were seen as being of ecological importance only in marginal and extreme environments. Now they are being discovered playing mainstream roles. Schleper's archaea findings^{5,6} are already causing scientists to reexamine the nitrogen cycle — the conversion of nitrogen to and from forms needed by plants and animals.

The success of studies such as Schleper's will not only spur work on as-yet-uncultured organisms, but also generate hypotheses about community dynamics and functions. Schleper notes that soil contains a lot of archaea, but that in moderate environments they are much less diverse than the bacteria they live alongside. So, did they always inhabit moderate environments, being overrun at a later point by more diverse bacteria? Or did they invade, and if so, how did they acquire their niche?

Nitty gritty

"We need more studies of the genes discovered by functional screens" in order to take such work further, says Rolf Daniel, a microbiologist at the Georg-August University in Göttingen, Germany. "Massive sequencing efforts are the starting point." Instead of cloning genes into bacteria and then studying the proteins that are expressed, large-scale sequencing allows scientists to analyse the genes directly, comparing their sequences with those of genes of known function. Ecology groups in the United States and Europe are trying to piece together the funding necessary for such sequencing capacity.

This is where Venter comes in. He plans to use the 'shotgun' sequencing technique — in which all the DNA in a sample is cut into small fragments, sequenced and then pieced together — to create a genetic inventory of the planet. In 2004, he produced a paper that sequenced the DNA in a sample of the Sargasso Sea: it yielded 1.2 million new genes and evidence for 148 new species⁷. In addition to the marine samples taken as part of his *Beagle*-emulating voyages aboard the *Sorcerer II*, Venter is also taking soil samples. Soil searchers have greeted this sally on to — indeed, into — their turf with a degree of scepticism and even jealousy. Only a handful of soil labs are funded well enough to construct metagenomic libraries like Schleper's, whereas the various institutes he has set up make Venter a one-man superpower in sequencing terms. Eugene Madsen, a microbial ecologist at Cornell University in Ithaca, New

AP PHOTO/M. HOUSTON

J. BURGESS/SPL

York, grudgingly agrees that Venter's enormous sequencing ability will probably be good for the research community. But he wonders whether Venter will end up with anything more than a hodgepodge of random sequences.

Venter freely admits that, compared with the clear waters of the Sargasso Sea, soils are proving a challenge. Sea water is well mixed and unstructured. You can take as much of it as you want — the Sargasso work took 200 litres — remove the DNA, sequence it, and come up with a pretty good picture of what was there and in what proportions. The diversity in soils means that sequencing a sample of more than a gram is overly ambitious, and this in turn means that the DNA from that gram needs to be amplified before it can be shotgunned. Venter concedes that this can create artefacts unless care is taken, and his Sargasso Sea work has been faulted for sequencing sample contaminants.

The whole truth

The focus of research need not be on single genes. Putting together individual microbial genomes from the jumble of library sequences is an obvious goal — albeit one that is as yet out of reach for samples that contain thousands of species. The most complicated system attempted so far was a community consisting of a handful of microbe species that lived in the extremely acidic waters draining from an old iron mine⁸. Even in this case, the genomes of just two species have been mostly completed, but the work is seen as a major step towards the ultimate goal of determining the ecological role of community members.

"Assembling genomes from the environment is a far more complicated task than the human genome," Venter says. Producing the sequence of a human's 3 billion base pairs required enough shotgunned snippets of DNA to cover the whole genome five to eight times

CULTURAL RENAISSANCE

Not surprisingly, traditionalists are concerned that classical microbiology's role will wane with the attention on metagenomics. Although culturing techniques have not proved effective for the majority of soil bacteria, microbes still need to be grown in the laboratory if one is to find out all that there is to know about them. Indeed, the need to culture these organisms

may now be more relevant than ever, given the increased interest in soil microbes that metagenomics is fostering.

Peter Janssen, a microbiologist at the University of Melbourne in Victoria, Australia, is leading a mini-renaissance of culturing organisms. One hope is that metagenomic information will provide clues to the lab conditions necessary to culture

elusive organisms. But at the moment Janssen's success is due more to patience.

By realizing that many organisms are inherently slow growing, he has single-handedly cultured hundreds of organisms, just by being willing to wait longer than his predecessors did. Sadly, he has made only a small dent in the biodiversity yet to be tapped. **V.G.**

over. "The soil work is like trying to assemble 40,000 human genomes simultaneously with only one-half to one-quarter of the sequencing coverage." Handelsman estimates that getting enough coverage to assemble the genomes of the species in 1 gram of soil would mean sequencing 250 billion base pairs; Venter shotgunned just 1 billion base pairs for the Sargasso Sea paper.

Schleper has constructed three libraries in total, covering 3 billion base pairs of DNA from just two different soil samples, a sandy ecosystem and a mixed forest soil, but hasn't sequenced them. She welcomes help doing the high-throughput work. "The interesting biology is interpreting and looking through it," she notes.

Even with all the sequencing power that money can buy, stitching together genomes of thousands of species is difficult owing to the plentiful horizontal transfer of genes. According to Jonathan Eisen, a genomicist at The Institute for Genome Research in Rockville, Maryland, the largest number of complete genomes so far assembled from a mixed DNA sample is just two: a study he was involved in managed to disentangle the genomes of a *Wolbachia* symbiont and its *Drosophila*

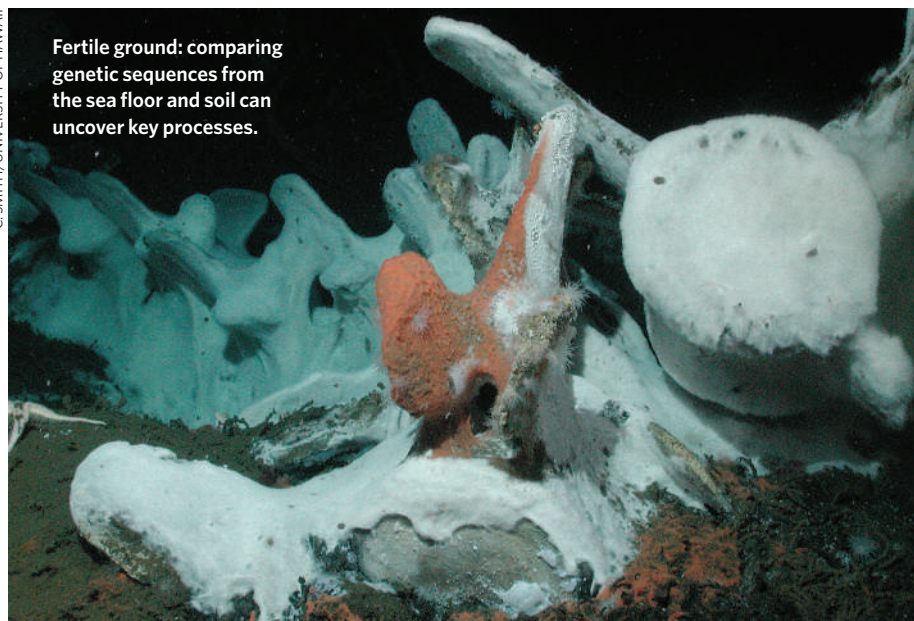
melanogaster host. Eisen stresses that even that was not easy.

To Schleper and others, however, the assembly of whole genomes is not necessarily an appropriate goal. Given the enormous amount of genetic exchange among microbes, treating the soil itself as an organism may be more valid. This is the approach used in a recent collaboration between Diversa and the US Department of Energy's Joint Genome Institute at Walnut Creek, California. The researchers put fragmentary sequences from their data into categories or 'bins' based on the metabolism they represent, rather than the gene or organism⁹. They compared the microbial communities in an agricultural soil with samples from sea water and with microbes from ocean-floor sediments recovered from sites where dead whales have decomposed. The comparisons showed that even incomplete DNA could characterize an environment and suggest ecological roles within it.

Whether or not individual genomes are pieced together, there is no doubt that the sequencing of soil genomes will be fertile ground for hypotheses about how microbial communities drive the ecological processes of a region. At the moment, says George Kowalchuk, a microbial ecologist at the Netherlands Institute of Ecology in Heteren, "even if you have the entire community genome, you're still far from predicting how it works." But at least we are heading in the right direction.

"I think this is an area that will have far greater impact on science than sequencing the human genome," says Venter. Not everyone might want to go quite that far, but Schleper is happy to insist: "This is really a new era in microbial ecology."

Virginia Gewin is a freelance science writer based in Portland, Oregon.



Fertile ground: comparing genetic sequences from the sea floor and soil can uncover key processes.

C. SMITH/UNIVERSITY OF HAWAII

1. Gillespie, D. E. et al. *Appl. Environ. Microbiol.* **68**, 4301–4306 (2002).
2. Wang, G. Y. et al. *Org. Lett.* **2**, 2401–2404 (2000).
3. D'Costa, V. M., McGrann, K. M., Hughes, D. W. & Wright, G. D. *Science* **311**, 374–377 (2006).
4. Sabehi, G. et al. *PLoS Biol.* **3**, e273 (2005).
5. Quaiser, A. et al. *Environ. Microbiol.* **4**, 603–611 (2002).
6. Treusch, A. H. et al. *Environ. Microbiol.* **7**, 1985–1995 (2005).
7. Venter, J. C. et al. *Science* **304**, 66–74 (2004).
8. Tyson, G. W. et al. *Nature* **428**, 37–43 (2004).
9. Tringe, S. G. et al. *Science* **308**, 554–557 (2005).

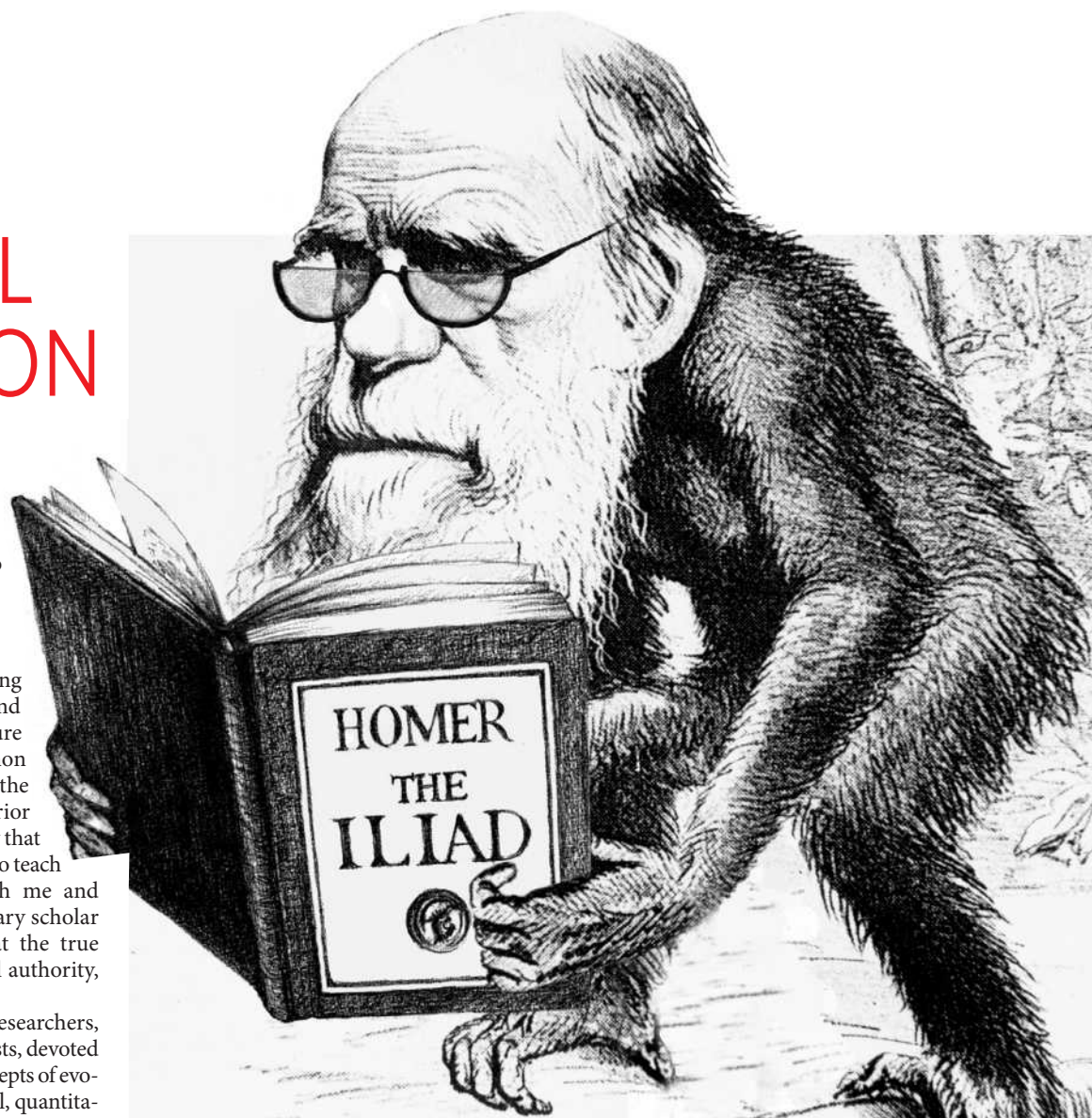
TEXTUAL SELECTION

Can reading the classics through Charles Darwin's spectacles reawaken literary study?
John Whitfield reports.

When, at the beginning of *The Iliad* — and Western literature — King Agamemnon steals Achilles' slave-girl, Briseis, the king tells the world's greatest warrior that he is doing so "to let you know that I am more powerful than you, and to teach others not to bandy words with me and openly defy their king"¹. But literary scholar Jonathan Gottschall believes that the true focus of Homer's epic is not royal authority, but royal genes.

Gottschall is one of a group of researchers, calling themselves literary darwinists, devoted to studying literature using the concepts of evolutionary biology and the empirical, quantitative methods of the sciences. "Women in Homer are not a proxy for status and honour," says Gottschall. "At bottom, the men in the stories are motivated by reproductive concerns. Every homeric raid involves killing the men and abducting the women." The violent world of the epics, he says, reflects a society where men fought for scarce mates and chieftains had access to as many women as slaves and concubines². And he thinks that everything written since Homer is open to similar analysis.

Literary darwinism is a mode of analysis; it's also a bit of a crusade, an attempt to shake up literary criticism. "Literary theory requires a theory of human nature, because literature is shaped by human motives and cognitive biases," says Joseph Carroll of the University of Missouri, St Louis. The problem, say the literary darwinists, is that for the past few decades the humanities have, in the case of critics deconstructing texts, denied the need for a theory of human nature, asserting that the study of texts can be concerned with nothing outside those texts. Or else they have been stuck on theories of human nature that are rooted in the subjective and the social.



Those influenced by freudianism, for example, might read a novel looking for hints of a child's sexual desire for its parent. A marxist would seek out economic and class conflicts. Carroll has no truck with this: "The theories up to this point have all had a little bit of the truth, but have also all been fundamentally flawed," he says. "None comes to terms with the fundamental facts of human evolution."

Literary darwinists believe that literature reflects a universal human nature shaped by natural selection, and as a result, read texts in terms of animal concerns such as mate choice, relations between kin, and social hierarchies. Such a scientific approach can meet with hostility. "At one meeting of the Modern Languages Association, someone stood up and called me a proto-fascist," says Nancy Easterlin, an expert in Romantic literature at the University of New Orleans, Louisiana, who uses ideas from cognitive science in her analysis of the mother-child bond in William Wordsworth's *Prelude*.

The tide may be turning, however. "The ideological resistance is crumbling pretty fast," says

the British author Ian McEwan, who has used scientific ideas in several of his novels. "Now things are spoken of that would have routinely got you called a Nazi a few years ago." The English department at Texas A&M University, in College Station, has recently approved a seminar on literary darwinism — the first university course on the subject, says Brett Cooke, the course leader and an expert in Russian literature.

Man to beast

So what does it mean to read literature through a darwinian lens? At one level, it can seem remarkably obvious. In their recent book *Madame Bovary's Ovaries*³, evolutionary psychologists David and Nanelle Barash argue that a darwinian understanding of female mate choice shows why the eponymous adulteress takes lovers who are more attractive and accomplished than her mediocre husband. This may sound crass, but Carroll argues that the approach is capable of subtlety. A darwinian analysis of Jane Austen's *Pride and Prejudice*, he says, goes beyond the simple idea that

SPR/N. SPENCER

women look for fortune in men, to show how such animal concerns are filtered through the vast flexibility of human behaviour, cultural conditions and individual variation.

"I don't look at *Pride and Prejudice* and try to sort out what is biological and what is cultural," says Carroll. "I look at it and examine the way underlying biological dispositions are organized in a specific cultural ecology. Nobody in the novel escapes the problems of mate selection, status and forming alliances. But the characters also integrate these concerns with human qualities, such as intelligence, character, morals and cultivation." The noble, romantic characters, such as Elizabeth Bennett and Darcy, integrate successfully, hiding their reproductive issues beneath their social graces. The more comic characters, such as Elizabeth Bennett's mother, do not (although in marrying off her daughters, she is quite the evolutionary success).

Romantic comedies play upon the audience's pleasure at seeing reproductive strategies rewarded; tragedies appeal by invoking recoil from maladaptive acts. "Stories that focus on non-normative behaviour, such as when Medea kills her children, take their punch from the audience's understanding that this is not how humans behave," says Gottschall.

Not everyone in the movement is equally keen on reductions to a purportedly universal human nature. "I've always had a love-hate relationship with evolutionary psychology. It's very interesting as far as it goes, but it marginalizes culture and other open-ended processes," comments David Sloan Wilson, a biologist at Binghamton University, New York. Wilson is also the editor, with Gottschall, of *The Literary Animal*⁴, a recent collection of essays on literary darwinism. But, Wilson adds, literature is an immense source of data on human behaviour: "It's the natural history of our species."

Anything that wakes literary study up to the idea of a shared human nature, reflected throughout literature, is to be welcomed, says McEwan, one of whose lectures is reprinted in *The Literary Animal*. "To think in evolutionary terms about human nature has helped me as a novelist, and to some extent as a reader," he says. An evolutionary emphasis might also help the study of literature to reverse its journey into obscurantism and irrelevance. "It's a tragedy, the way that literary criticism has lost its place," he says. "I don't read much literary theory, especially of the kind that has dominated the academy for the past few decades: there's been a great flatus of nonsense and pseudoscience."

But even if literary theory is starting to get over its objection to evolution, it may be more



FOX-WARNER

Achilles (Brad Pitt) ponders the mother-child bond before heading off for some mate selection in *Troy*.

resistant to the other main item on the darwinist's agenda — quantitative research methods. "Among literary folk, the fear of quantification is greater than the fear of evolution," says Wilson. Gottschall agrees: "All literary scholars think they're mathematically disabled."

A common tale

Gottschall has analysed a database of folk tales from around the world to test the idea that a focus on beautiful princesses in need of rescuing and dashing hero is not just a product of patriarchal attitudes in European societies, as some feminist critics have claimed. He found that all around the world, the majority of folk tales feature brave heroes marrying beautiful heroines, with the two living happily ever after⁵.

Gottschall and Carroll, collaborating with psychologists, are also currently analysing the data from an online questionnaire that gathers people's responses to characters in nineteenth-century fiction; they aim to see how these compare to the personality categories and goals defined in evolutionary psychology⁶.

By borrowing the scientific method, says Gottschall, literary scholars can work out what a story is 'really' about, not in some ultimate, metaphysical sense, but in the sense of whether a wide range of people interpret a work in the same way. Such an approach, he says, is needed if literary scholarship is to create testable, durable knowledge — and to prevent arguments being settled solely by who deploys the sharpest rhetoric and the best memory.

"Literary critics are looking for patterns of meaning, rather than trying to produce an overarching theory of life." — David Amigoni

David Amigoni, a specialist in Victorian prose at Keele University, UK, agrees that there can be value in darwinian interpretations, as well as in reading Darwin. But he insists that attitudes and readings are mutable. "Proving claims of truth is not necessarily what literary critics are looking to do. They're looking for patterns of meaning, rather than trying to produce an overarching theory of life." He doubts that graphs and statistics can say much about literature: "The emphasis on hard data will probably be a bit strange to a lot of literary critics. I have a concern that something that ends up in numbers hasn't really taken account of literary value."

Gottschall, though, wants to move beyond literary value — or for that matter, traditional literary criticism. Literary scholars may adopt their theories from other branches of knowledge, but they also push them outwards, using their theoretical frameworks to analyse philosophy, science, history and gender politics, for example. Ultimately, the theories of human nature that become widely held in a society will influence how that society believes people respond to their environments, and how they should be treated. "Literary scholars aren't harmless," Gottschall says. "When we get it wrong it matters."

John Whitfield is a freelance science writer based in London.

1. Homer *The Iliad* (Penguin, London, 1950).
2. Gottschall, J. *The Rape of Troy: Evolution, Violence, and the World of Homer* (Cambridge Univ. Press, in the press).
3. Barash, D. P. & Barash, N. R. *Madame Bovary's Ovaries: A Darwinian Look at Literature* (Delacorte, New York, 2005).
4. Gottschall, J. & Wilson, D. S. (eds) *The Literary Animal: Evolution and the Nature of Narrative* (Northwestern Univ. Press, Evanston, Illinois, 2005).
5. Gottschall, J. in *The Literary Animal: Evolution and the Nature of Narrative*. (eds Gottschall, J. & Wilson, D. S.) 199–224 (Northwestern Univ. Press, Evanston, Illinois, 2005).
6. <http://survey.ehap.isr.umich.edu/carroll-intro>

BUSINESS

The right combination

Instead of the painstaking process of developing new drugs, one Boston-based company is making its mark by pairing up those we already have. **Meredith Wadman reports.**

Drug research largely boils down to the search for 'magic bullets' — molecules that act against single, specific biological targets. But putting each candidate bullet through its paces is a laborious and expensive business. Flying in the face of convention, a young biotech company in Boston, Massachusetts, has adopted a very different approach, and has so far been well rewarded for its efforts.

CombinatoRx (pronounced Combinatorix) is the brainchild of Harvard chemistry PhD dropout Alexis Borisy and several pals, all of whom worked in the lab of leading chemical biologist Stuart Schreiber.

In 1999, over hot chocolate in a favourite cafe, Borisy, Michael Foley and Brent Stockwell dreamed up a scheme to abandon the search for new molecules in favour of redeploying existing drugs. They planned simply to fire combinations of existing drugs at a range of biological targets that mimic diseases to see which proved effective, and then patent and sell the successful pairings. Back-of-the-envelope maths told them that the 2,000 unique drugs already approved by the US Food and Drug Administration — most of which are off-patent — would yield some 2 million two-drug combinations to be tested.

It was a "very simple, dumb, empirical, brute-force approach" to drug discovery, says Borisy. But it made intuitive sense. Single drugs usually modify a particular cell signal involved in a disease. Yet the body rarely sends a cellular signal using just one molecule — instead several signalling networks are typically acted on by lots of molecules at multiple points. So rather than trying to knock out one point on a pathway, Borisy and his friends reckoned it might be possible to use two drugs to hit the right combination of targets.

Business-wise, this search for molecular synergy had a lot in its favour. By testing approved drugs — whose chemistry and pharmacology are already understood — the company could, they hoped, catapult itself through the early stages of drug discovery, saving tens of millions of dollars.

"The discovery process was reduced to weeks, rather than years," says Julian Adams,

chief scientific officer at Infinity Pharmaceuticals in nearby Cambridge. "That's the coolest part of the whole enterprise."

Within nine months, the trio — joined by Curtis Keith, also from Schreiber's lab — had secured seed funding and hired their first employees. They began building a prototype screening system in a basement lab in downtown Boston. That involved adapting existing systems to accommodate pairs of compounds, instead of individual agents. They ended up with a proprietary system that evaluates 36 different dosing configurations at once for many pairs of compounds. The first million or so combinations they screened produced 300 interesting pairs. Seven of these are now in clinical trials, with a much larger group in the pre-clinical pipeline, according to Borisy, who is now the company's chief executive.

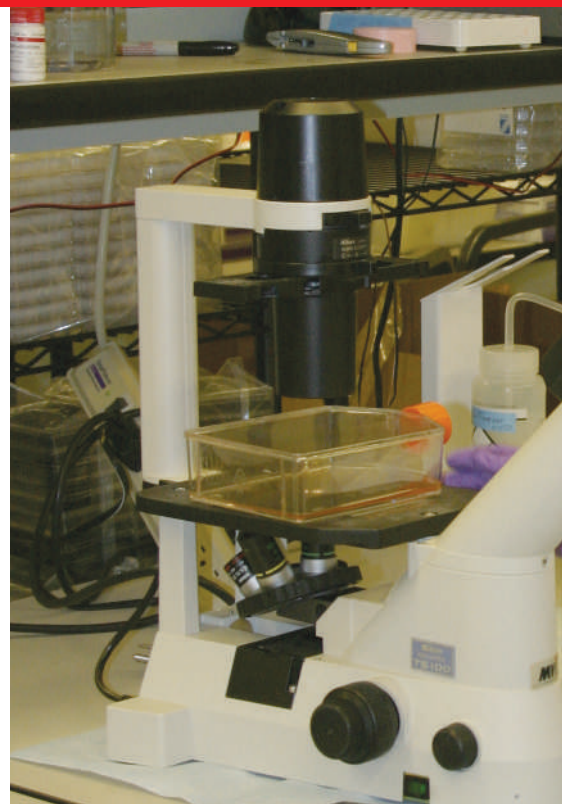
In their spare time, the group raised \$90 million in venture capital (most of it now spent) and signed licensing deals with other companies worth \$58 million. CombinatoRx went public in November, adding another \$44 million to its coffers (see chart).

"Most people would have said: 'This is crazy, these four kids right out of grad school starting a publicly traded company,'" muses Stockwell, a co-founder and adviser to the company, who is now at Columbia University in New York. "But here we are with \$100 million in the bank."

The current phase I and II clinical trials are testing combinations against rheumatoid arthritis, psoriasis, asthma and metastatic cancers such as colon and lung. The mixtures involved are nothing if not unconventional. For one, the anticlotting agent dipyridamole is

"The company is doing something really unique."
— Danny Cummings

CombinatoRX share price



Joint effort: a CombinatoRx researcher tests a drug combination on cancerous cells.

paired with the steroid prednisolone to treat rheumatoid arthritis. The idea is that the dipyridamole boosts the effect of the steroid, allowing a lower dose of prednisolone and so causing fewer side effects. In another, the anti-fungal pentamidine is battling cancer aided by the antipsychotic chlorpromazine.

Dynamic duos

"They're doing something really unique," says Danny Cummings, an analyst who reports on CombinatoRx for Hoover's, a business information service in Austin, Texas. "In five years, they've got seven products into clinical trials for a total cost of less than \$50 million. The average cost to develop a new prescription drug is \$802 million."

But some observers have expressed concerns about safety, arguing that uncharted toxicities could emerge just as easily as unpredicted therapeutic effects. Borisy counters that the company's safety testing is thorough. "Synergy is rare, whether good or bad," he adds. "We will certainly be safer than any new chemical entity."

Nevertheless, investors in the company face two big questions. What will stop doctors from writing separate prescriptions for cheap, generic versions of the components that make up the products? And will US health insurers pay out for a combination when the generics are each available more cheaply?

"It's the most obvious concern," says Adams. "The only way I can think around it is that they are using different doses of the drugs than are currently manufactured, making a combination drug more convenient."

Borisy says the company has done that, and



COMBINATORX

more besides, to protect its carefully patented combinations. The company is, for instance, using timed-release formulations that make its two-part recipes difficult to mimic.

His arguments have certainly swayed some investors. "They've gone to great lengths to refine the timing of the combinations and the dosage," says William Hunter, chief executive of Angiotech Pharmaceuticals, a biotech firm in Vancouver, Canada. Last October, Angiotech agreed to pay CombinatoRx \$42 million — plus up to \$460 million in downstream milestone and royalty payments — for access to its database and the rights to develop active combinations for specific applications, such as preventing internal scarring after surgery.

Simple considerations of convenience may also play a big role, argues Ted Ashburn, head of business development at Dynogen Pharmaceuticals in Waltham, Massachusetts. "It is inherently easier for doctors to write, and for insurers to reimburse, one prescription versus two — and for patients to take a single tablet versus two."

One precedent suggests that CombinatoRx will be able to charge reasonable, if not exorbitant, rates for its pills. NitroMed, of Lexington, Massachusetts, makes BiDil, a congestive heart-failure drug targeted at African-Americans. Based on two off-patent components, BiDil costs up to \$11 per day — roughly double the cost of the generics.

That, the CombinatoRx crew hopes, is a harbinger of things to come. "I obviously have a personal, financial and emotional stake in the company," says Stockwell. "But if you really believe in the approach, then CombinatoRx should by rights become the largest and most successful pharmaceutical company in the history of the industry." ■

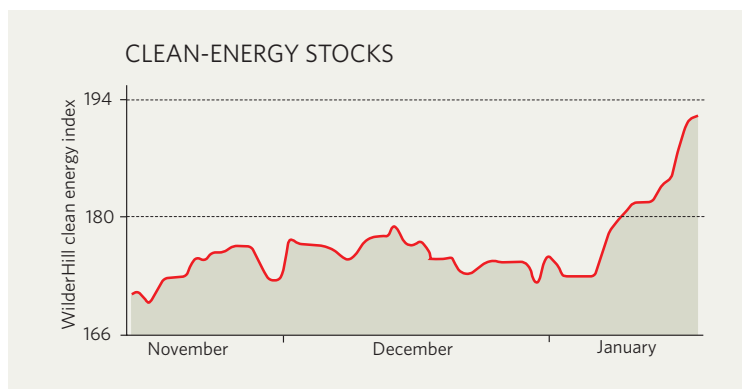
IN BRIEF

BIOCRYST BUOYED The US Food and Drug Administration last week agreed that potential flu drug peramivir can be fast-tracked through the approval process. Made by BioCryst Pharmaceuticals of Birmingham, Alabama, peramivir is an injectable drug that may be able to treat avian flu. News of its fast-track status sent BioCryst's share price soaring — it closed up 17.6%, at \$22.41, on the day of the announcement. The fast-track process means that fewer data have to be shown before the drug's approval for use in humans. BioCryst's shares have more than doubled in value since the start of October, boosted not only by peramivir but also by the launch of an anticancer drug trial and a licensing deal with Roche (see *Nature* 438, 737; 2005).

CHIPPING AWAY Just one day after Intel disappointed Wall Street with fourth-quarter results that were weaker than expected, its chief competitor revealed record sales for the same period. On 18 January, Advanced Micro Devices of Sunnyvale, California, the second-largest manufacturer of the chips that run personal computers, announced sales of \$1.84 billion for the fourth quarter, up 45% on the previous year. This suggests that despite being the world's leading semiconductor maker, Intel may be losing more market share to its arch-rival than it has acknowledged.

PFIZER STRUGGLES Hounded by generic competition, Pfizer, the world's largest drug-maker, reported lacklustre results for the fourth quarter of 2005. At \$2.73 billion, net income was down 3.3% on the previous year. Fourth-quarter revenues fell 8.9% from 2004, to \$13.59 billion. In a statement, Henry McKinnell, Pfizer's chief executive, called 2005 "one of the most difficult years in memory". But the latest results weren't as bad as investors had feared, and the firm's share price closed up 4% at \$24.97 on 19 January, the day the results were announced.

MARKET WATCH



Stocks in companies with interests in alternative energy sources breezed into the new year, with investors particularly excited by the prospects for solar power.

The WilderHill Clean Energy Index (ECO on the American Stock Exchange) moved forward on the back of news about successful share offerings in a couple of solar-panel companies. The sector is also benefiting from a general sense that the uncertainty about oil and gas supplies that dogged major nations in 2005 won't dissipate any time soon — leaving people more positive about energy sources that were once considered to be on the fringe.

"Solar stocks are taking off," says Robert Wilder, whose company, WilderShares in Encinitas, California, maintains the ECO index. An initial public offering (IPO) for SunPower, a solar-panel maker spun off by California-based Cypress

Semiconductor in November, raised \$139 million. Then last month, an IPO for China's Suntech Power netted \$455 million — not bad at a time when many young companies are struggling to get initial public offerings off the ground. "Solar-panel companies are growing at 30–40% each year, and they are typically selling everything they can produce," Wilder says.

But it's not all rosy for alternative-energy suppliers. Companies working with hydrogen or fuel cells "continue to disappoint", Wilder says. Many of the smaller ones, he adds, are now "running out of money".

This week, WilderHill and London-based New Energy Finance will launch a new market index tracking more non-US companies than ECO, reflecting the strength of demand for energy from renewable sources in Europe and Asia. ■

Colin Macilwain

SOURCE: ECO

Biodiversity data are out of local taxonomists' reach

SIR — Exchange of information about biodiversity is mandated by the legally binding international Convention on Biological Diversity, as are monitoring and benefit sharing. Yet researchers in the developing world, where most of the biodiversity is found, are unable to access much of this information. This impedes the monitoring of biodiversity: monitoring depends on the proper identification of species, and this is hindered by a lack of both specialists and access to relevant taxonomic information.

The number of online publications with taxonomic content is increasing, and online tools are becoming available to mash up taxonomic with other information, for example at ispecies.org (see "Mashups mix data into global service" *Nature* **439**, 6–7; 2006). But copyright and high costs put this information beyond the reach of many in the developing world — which is home to more than 95% of species whose descriptions have been published. More than half the 1,600 descriptions of new ant species published in the past ten years are copyrighted, for example, but none are in journals published in the developing world (see antbase.org).

This seems little better than biopiracy: taking biodiversity material from the developing world for profit, without sharing benefit or providing the people who live there with access to this crucial information.

A simple solution would be to treat species descriptions as we do gene sequences, and have them openly accessible. Open-access descriptions of new species could then be a mandatory factor in making them valid under the various codes of biological nomenclature. A recent Commentary by Andrew Polaszek and colleagues ("A universal register for animal names" *Nature* **437**, 477; 2005) describes how the International Commission on Zoological Nomenclature proposes to facilitate this process for animal descriptions, through a register called ZooBank. However, present copyright laws prevent the mandatory inclusion of what would be an immensely useful piece of information, the actual description of the species.

Donat Agosti

American Museum of Natural History, New York, New York 10024-5192, USA

No clear evidence to disprove optics thesis

SIR — David G. Stork, in his review of the book *Optics, Instruments and Painting, 1420–1720: Reflections on the Hockney–Falco Thesis* ("Tracing the history of art" *Nature*

438, 916–917; 2005) reports the claim that appropriate concave mirrors to project optical images onto a canvas for tracing could not have existed in the fifteenth century. He concludes that the book may "close the door on the Hockney–Falco tracing thesis".

But our thesis is about the use of optics, not necessarily concave mirrors (although there is strong circumstantial evidence for mirrors). David Hockney and I have explicitly written during the past five years that none of the optical evidence we have found allows us to distinguish between the use of refractive versus reflective optics; either or both types are possibilities.

The spectacles and the magnifying glass in Tommaso da Modena's 1351 frescos would each have been able to project appropriate images, as would the spectacles in van Eyck's van der Paele altarpiece of about 1435.

Charles M. Falco

College of Optical Sciences, G-S Building 77, University of Arizona, Tucson, Arizona 85721-0077, USA

Branding can be justified in vital conservation research

SIR — Your News Story "Animal-rights group sues over 'disturbing' work on sea lions" (*Nature* **436**, 315; 2005) shows how wildlife researchers face escalating criticism over their use of invasive procedures to collect ecological data for population management. Although we appreciate and agree that all due consideration should be paid to the welfare of study animals, we argue that invasive research techniques are sometimes inevitable and integral parts of animal research in the global pursuit of biodiversity conservation.

The branding of seals is one technique criticized for being an invasive conservation tool. The technique has, however, provided important insights on juvenile survival that help to explain the processes involved in recent population declines — and a long-term programme of branding southern elephant seals (*Mirounga leonina*) on Macquarie Island, Australia, has shown that it has no discernable long-term impact on survival or condition (C. R. McMahon *et al.* *J. Wildl. Mgmt.*, in the press).

The issue of branding on Macquarie Island became so controversial that the programme was suspended indefinitely (J. A. Jabour-Green and C. J. A. Bradshaw *J. Nat. Conserv.* **12**, 25–39; 2004). Similar intervention has occurred in conservation research on two endangered species: New Zealand sea lions (*Phocarcos hookeri*) in New Zealand and Steller sea lions (*Eumetopias jubatus*) in the United States. Such political interference on the basis of animal-welfare issues is difficult to justify, considering that, in the United States alone, some 36 million beef cattle are hot- or

cryo-branded each year, according to the US Department of Agriculture.

Effective conservation of species is urgently needed in the current biodiversity crisis. To this end, we urge scientists and welfare lobbyists to combine their efforts and consider objectively the pertinent issues balancing the desperate need for sound biological information with animal welfare.

C. R. McMahon*, C. J. A. Bradshaw†, G. C. Hays*

*Institute of Environmental Sustainability, School of Biological Sciences, University of Wales

Swansea, Singleton Park, Swansea SA2 8PP, UK

†School for Environmental Research, Institute of Advanced Studies, Charles Darwin University, Darwin, Northern Territory 0909, Australia

Don't forget randomness is still just a hypothesis

SIR — Anton Zeilinger's bold essay "The message of the quantum" (*Nature* **438**, 743; 2005) claims "the discovery that individual events are irreducibly random is probably one of the most significant findings of the twentieth century." But we should not forget that the claim of true randomness has not yet been backed by evidence. Neither Heisenberg's uncertainty principle nor Bell's inequality exclude the possibility, however small, that the Universe, including all observers inhabiting it, is in principle computable by a completely deterministic computer program, as first suggested by computer pioneer Konrad Zuse in 1967 (*Elektron. Datenverarb.* **8**, 336–344; 1967).

The principle of Occam's razor, which is fundamental to theory-building, favours simple explanations (describable by few bits of information) over complex ones. But if the Universe's history really included many truly random events, an enormous amount of information would be necessary to describe all the random observations inexplicable by the known, simple, elegant, compactly describable laws of physics.

A few previous attempts at discovering a pseudo-random generator behind seemingly random physical events have failed (see T. Erber and S. Putterman *Nature* **318**, 41–43; 1985). But as long as the randomness hypothesis has not been verified, physicists should keep trying to falsify it and search not only for statistical laws but also for deterministic rules explaining any type of hitherto unexplained apparent randomness.

Juergen Schmidhuber

IDSIA, Galleria 2, 6928 Manno-Lugano, Switzerland and

Robotics and Embedded Systems, Technische Universität München, Boltzmannstraße 3, 85748 Garching bei München, Germany

Contributions to Correspondence may be submitted to corres@nature.com.

BOOKS & ARTS

Agents of destruction

An in-depth look at the state of biological-weapons programmes across the world.

Deadly Cultures: Biological Weapons Since 1945

edited by Mark Wheelis, Lajos Rózsa & Malcolm Dando
Harvard University Press: 2006. 479 pp.
\$59.95, £37.95

Jens H. Kuhn

Biological weapons have received considerable attention in the media and in the scientific community since the terrorist attacks of 11 September 2001 and the mailing in the United States of letters laden with *Bacillus anthracis* spores. Today, the debate about bioweapons is often characterized by unscientific prophecies and unlikely doomsday scenarios, a lack of discrimination between biowarfare, bioterrorism and biocrime, and a lack of uniformly accepted definitions for the terms biosafety, biosecurity and biodefence.

The threat of biological attacks has existed for many decades. Germany, for instance, established a biological sabotage programme as long ago as 1915 with the aim of incapacitating or killing military livestock of the Allied forces. Other countries, including Britain, Canada, France, Japan, the United States and the Soviet Union, initiated bioweapons research and development (R&D) programmes in the 1920s and 30s to develop weapons targeting animal, human or plant populations.

Historians and political scientists are well aware of these events. One of the most compelling and comprehensive scholarly books on the history and goals of these programmes, *Biological and Toxin Weapons: Research, Development and Use from the Middle Ages to 1945*, edited by Erhard Geissler and John Ellis van Courtland Moon, was published in 1999 by the Stockholm International Peace Research Institute (SIPRI). Scientific treatises on offensive programmes of individual nations after 1945 are also available, but a book contrasting all the known programmes in one volume has been conspicuously absent. *Deadly Cultures* aims to fill this gap. In fact, the editors view the book as a sequel to the acclaimed SIPRI volume.

The first nine chapters describe the history of bioweapons programmes in Canada, France, Iraq, the non-Soviet Warsaw Pact countries, South Africa, Britain, the United States and the Soviet Union. The next chapters describe national efforts to develop weapons aimed at killing crops and animals, as well as 'non-lethal'

weapons, and there are detailed analyses of alleged biological attacks. Further chapters describe legislation aimed at preventing the proliferation of bioweapons. Finally, to address the concerns of recent years, there is an analysis of biological terrorism, and a comparison of it with national biowarfare initiatives.

Deadly Cultures proves to be a worthy successor to the SIPRI volume. The authors are renowned experts from various countries. Most of the chapters draw mainly on primary sources, many of which are recently declassified documents. Individual chapters are not mere summaries of more elaborate books on a particular programme published by the same authors, but glimpses into ongoing research with new information. References to popular science books, unverified personal accounts of former bioweaponeers or defectors, and newspaper articles only appear when sufficient access to primary sources was not yet possible.

What do we learn from the book? Before 1945, several countries started biowarfare programmes because of a perceived but essentially non-existent biological threat from Nazi Germany. In a similar way, bioweapons research surged in Canada, Britain and the United States shortly after 1945 because of a perceived threat from the Soviet bioweapons programme. The West's activities did not go unnoticed by Soviet intelligence, however, and eventually the Soviet Union transformed

its limited activities into the world's largest bioweapons R&D programme, partly in response to the growing threat from the West.

Learning from this, van Courtland Moon concludes that current activities at the US Department of Homeland Security, which might challenge the international conventions that prohibit bioweapons development, could in effect "give the United States a modern offensive [bioweapons] capability", at least in the eyes of nations not friendly with it. Martin Furmanski and Mark Wheelis agree that current biodefence activities could "provoke reactions that reduce international security". In a final analysis chapter, the editors, along with Graham Pearson and Julian Robinson, emphasize that states "would be wise to review their biological defense programs from the point of view of other states, which might form incorrect perceptions" and begin offensive programmes based on incorrect threat assessments.

Deadly Cultures explains that in the years after 1945, military planners on all sides viewed bioweapons to be equal in their destructive potential to nuclear devices. However, within the next 20 years, extensive field trials with simulants on often unsuspecting civilians demonstrated that bioweapons could hardly reach the devastating effects of actual weapons of mass destruction. Today, one often hears about the threat of biological weapons of mass destruction again. This notion is based

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Dangerous delivery: the anthrax scare from spores being mailed in the post raised fears of bioterrorism.

FBI/REUTERS

on the idea that such weapons could now be developed because modern methods would overcome previously insurmountable technical obstacles. But by describing how difficult and costly it was to develop any functional bioweapons before the 1990s, the book could emphasize the false nature of the frequently repeated idea of bioweapons as a 'poor man's atom bomb'.

We also learn that the idea of terrorist 'sleeper cells' goes back to 1951, when the CIA suspected that Soviet saboteurs were living unnoticed on US soil while waiting for an order from Moscow to launch a biological attack. Reading on, I was surprised to learn that Hungary probably undertook some offensive bioweapons activities before 1945, and that this programme might have had ties with a clandestine, yet-to-be-described Italian programme. The equally new revelation that Czechoslovakia might have kept stocks of variola virus up to 1994 is, if proved correct, stunning and frightening. Smallpox (the disease caused by this agent) was eradicated in the late 1970s, and official stocks of variola virus have since been permitted in only two laboratories, one in the United States and one in Russia.

The chapter on bioweapons R&D in Iraq provides a fascinating comparison of official Iraqi statements with the actual findings of the United Nations Special Commission, the UN Monitoring, Verification and Inspection Commission, and the Iraq Survey Group. Once again it is emphasized that no evidence of bioweapons development was found in Iraq after 1996, that the UN monitoring regime was very effective, and that its negative findings were correct.

The description of South Africa's bioweapons programme, which began in 1981, shows that only 'crowd control' and assassination weapons, rather than weapons of mass destruction, were developed, and that the programme was not using sophisticated molecular biology. In discussing the allegations of bioweapons use by the United States in Korea and China in 1952, and by the Soviet Union during the war in Afghanistan, Furmanski and Wheelis take into account most available data on these allegations from either side so that assessment could not be misconstrued as parochialism or ill-conceived patriotism. They conclude that most allegations are probably false.

Deadly Cultures is written eloquently and has been edited superbly. The chapters have a uniform style and organization; scientific and political terminology is used in a consistent and correct manner throughout; and abbreviations are used only where absolutely necessary. In contrast to most other books on bioweapons, the editors have almost always used up-to-date taxonomy of biological agents, as well as the differentiation of agents and the diseases they cause. The authors also included the original names of all institutes involved in bioweapons R&D. This is not a trivial point as French, Iraqi or Russian institute designations

have been translated differently in the past, and were also frequently changed during decades of reorganization, confusing both analysts and interested laymen.

I wish the book contained more references to biological anti-material weapons. Research activities on, for instance, rust-inducing, oil-degrading or asphalt-destroying agents are increasing and possibly challenge the 1972 Biological and Toxin Weapons Convention. Also, a chapter analysing the suspected motives (or lack thereof) of 'bioterrorists' would have been helpful, as it is by no means clear that terrorist organizations actually consider using

weapons that would target politicians, civilians and themselves alike. Finally, in the chapter on legal constraints on bioweapons, there was no reference to the Convention on the Prohibition of Military or Any Other Hostile Use of Environmental Modification Techniques. These minor concerns aside, however, *Deadly Cultures* is informative, meticulously researched, important in its message, and a fabulous read for both scholars and interested scientists. ■

Jens H. Kuhn is in the Department of Microbiology and Molecular Genetics, Harvard Medical School, Southborough, Massachusetts 01772, USA.

What is it like to speed date?

Conversations on Consciousness

by Susan Blackmore

Oxford University Press: 2005. 288 pp.

£18.99, \$23

Adina Roskies

The hard kernel of the mind-body problem — how we get first-person experience out of a purely physical object like the brain — was famously articulated by Thomas Nagel in a paper entitled 'What Is It Like To Be A Bat?' (*Phil. Rev.* **83**, 435–450; 1974). The question of 'What is it like?' concerns the phenomenality and subjectivity of experience, and has come to be known as the hard problem of consciousness. This is the central focus of Susan Blackmore's latest book, *Conversations on Consciousness*, a compendium of 20 interviews she conducted with major figures in the field of consciousness studies. The illustrious but motley crew includes philosophers of radically different stripe such as David Chalmers, Pat and Paul Churchland, Daniel Dennett and John Searle; psychologists V. S. Ramachandran, Kevin O'Regan and Daniel Wegner; neuroscientists Francis Crick and Christof Koch; and explorer of altered states Stephen LaBerge.

The experience of reading Blackmore's book is the intellectual analogue of what it must be like to participate in the popular institution of speed dating, that maximally efficient method of meeting a potential partner. Blackmore devotes roughly 13 pages to each interview transcript, which reads roughly like a 10- to 15-minute conversation. Just as you might expect when interviewing potential romantic partners, some encounters with Blackmore's interviewees leave you wanting more, whereas others fail to connect and would be excruciating but for their merciful brevity.

In the course of Blackmore's discussions about how subjective experience might result from the operation of the three-pound hunk of meat that is our brain, she explores her subjects' disagreements with others' theories, their views about free will, and their opinions about the value of meditation and Eastern religious practices (the intellectual equivalent, I take it, of "What's your sign?"). She also poses personal questions to the interviewees, such as why they were drawn to studying consciousness in the first place, and whether their work has influenced the way they approach their own personal experiences.

Researchers have often wondered how we can get first-person experience from the matter that makes up our bodies.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

H. SOCHUREK/CORBIS

The book best succeeds in providing a very brief survey of the multitude of positions occupied by thinkers in this area. The lack of agreement on any issue, such as whether there really is a hard problem — or if there is, what it is — is striking. Some theorists think that the problem is really hard. Even when we understand how the brain accomplishes its astounding variety of complex tasks, such as visual recognition, memory, planning and so on (the 'easy problems' of consciousness), something will be left unexplained: how or why these intelligent behaviours are accompanied in us by conscious states. Those who believe in the hard problem generally believe in the possibility of 'zombies' — beings who function exactly as we do, yet lack the mysterious spark of consciousness, so "all is dark inside". Others think the hard problem isn't really that hard, and that the problem of subjectivity will dissolve once we have a handle on the easy problems. Still others claim that the problem itself is illusory.

Because of the extremely light hand Blackmore takes in editing, the often quirky personalities and mannerisms of the interviewees shine through the text. The effect is magnified when you know the people: I could hear, for instance, Ned Block's enthusiastic voice and Crick's wry quips about philosophers in my mind's ear. This gives the book some added appeal: readers really get a sense of 'what it is like' to talk to these people. A few of the interviews with people I've never met made me wish I had a chance to explore their views further over dinner and a good bottle of wine, but others left me cold. Blackmore herself comes across as spunky and clever, and the probing follow-up questions she occasionally asks prevent the interviews from seeming too repetitive and boring.

However, if you are serious about meeting an intellectual soulmate in the quest to understand consciousness, speed dating may not be for you. The book is rather unsatisfying for anyone with a deep interest in the issues, for no position is articulated clearly enough for readers to see the depth of the problems or the breadth of knowledge (or ignorance) that characterizes our current understanding of issues related to consciousness. Despite Blackmore's obvious intelligence and familiarity with the issues, at crucial points she does not press her interviewees hard enough or deeply enough to provide us with truly novel insights.

Conversations on Consciousness provides an introduction to a variety of positions, but is too cursory to make possible their evaluation. For that, one would need to spend a few evenings alone with the works of one or another of the thinkers. Like speed dating, *Conversations on Consciousness* is low-risk, but ultimately also low-payoff. It is, at best, a good way to guide an interested novice into the field. Second date, anyone? ■

Adina Roskies is in the Department of Philosophy, Dartmouth College, Hanover, New Hampshire 03755, USA.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Flame and fortune? Antoine Laurent Lavoisier (left) beat Joseph Priestley to the discovery of oxygen.

Burning ambition

A World on Fire: A Heretic, an Aristocrat, and the Race to Discover Oxygen

by Joe Jackson

Viking: 2005. 384 pp. \$27.95, £17.99

Bernadette Bensaude-Vincent

One of the most famous episodes in the history of chemistry is the race for priority between the two rival champions of oxygen, Joseph Priestley and Antoine Laurent Lavoisier. Priestley was a Unitarian minister who divided his life between laboratory experiments and theology, and was forced to move from England to exile in the United States. Lavoisier was a young, ambitious and wealthy academician who never left France and met a tragic end in 1794, when he was guillotined by French revolutionaries. Joe Jackson plays nicely on the contrast between the two men in his extremely readable book *A World on Fire*. The title refers both to the role of oxygen in combustion, first established by Lavoisier, and to the context of scientific competition and political upheaval.

Jackson tells the story in the manner of a standard historical narrative, in chronological order, occasionally interrupted by glimpses of the broader cultural and political context. However, some of these interludes, such as the chapter on the guillotine, which speculates on how long its victims had to suffer before they died, do not seem particularly relevant. If the goal of the book was to weave together science and politics, it is not fully achieved. And this is not just because of the spelling mistakes and incorrect dates (for example, Descartes's *Discourse on Method* was published in 1637, not 1677).

The narrative fails to adequately recreate the scientific milieu of the late Enlightenment in Britain and France. Jackson consulted Priestley's archives, but he did not rely on primary sources for the French part of the story. He didn't even get his information from the recent wealth of scholarly publications on the

chemical revolution. For instance, Frederic L. Holmes' *Antoine Lavoisier: The Next Crucial Year* (Princeton University Press, 1998) would have been a useful source for describing the pathway to the discovery of oxygen, especially as it is based on a close examination of Lavoisier's laboratory notebooks of the year 1773. As a result, Jackson's book reinforces some old clichés, such as the view of Lavoisier's career as a systematic development of a seminal idea, a revolutionary plan meant to overthrow Georg Stahl's phlogiston theory.

More importantly, Jackson's early chapters suggest that pre-lavoisierian chemistry was an inconsistent, empirical science, clinging to the ancient doctrine of the four elements. In truth, historians of eighteenth-century chemistry describe a booming field, based on more robust notions: not only had the four elements been redefined in terms of simple substances and agents or instruments, but laboratory practices were guided by tables of affinities.

The narrative itself suffers a major bias, being written from a present-day perspective. Because Jackson knows that the 'dephlogisticated air' that Priestley released from mercury calx was oxygen, he doesn't create any dramatic suspense. He assumes from the beginning that Priestley was wrong and Lavoisier was right. It would have been more interesting to show how the identity of oxygen was constructed through the confrontation between Priestley and Lavoisier. The contrast between Lavoisier's academic experiments, using sophisticated and expensive instruments, and Priestley's attachment to more democratic and qualitative practices, was described in a more balanced way by Jan Golinski in *Science as Public Culture* (Cambridge University Press, 1992). Jackson portrays Priestley as a complex and interesting character, but makes no effort to understand his strong convictions and religious beliefs. In contrast, the two-volume biography by Robert Schofield,

The Enlightenment of Joseph Priestley and The Enlightened Joseph Priestley (Pennsylvania State University Press, 1977 and 2005), provides a detailed account of Priestley's multiple facets.

Popular historical narratives should not be blamed for distorting scholars' historical accounts; after all, each historical narrative is a reconstruction of the past, even those based on a detailed analysis of primary sources. Popular

historical accounts can convey a clear picture of the period and the characters, something achieved by the Open University video *The Publicity of Oxygen* (BBC, 1993). Some of them openly presented as fictions raise stimulating issues. This was the case with *Oxygen*, a play written by Carl Djerassi and Roald Hoffman that created a 'retro' Nobel to be awarded to the discoverer of oxygen. The discussions of

the Nobel committee as it decides between Carl Wilhelm Scheele, Priestley and Lavoisier prompt reflection about the mechanisms of discovery attributions. Historical fiction like this may be more useful and more pleasant than inaccurate pseudo-realistic accounts. ■ Bernadette Bensaude-Vincent is in the Department of Philosophy, Université Paris X, 200 avenue de la république, 92001 Nanterre, France.



KIM KYUNG-HOON/REUTERS

Stamping his authority

The Hwang scandal highlights the dangers of hyping science.

Martin Kemp

The repercussions of the falsification of stem-cell research by Korean scientist Woo Suk Hwang will reverberate around the scientific community for years to come. And the public dimension seems just as momentous, especially when we consider the mechanisms that elevated this former veterinarian, who was famous only in his home country in 1999, to become one of *Time* magazine's people of the year in 2004. Hwang's rise involved celebrity-minded scientists, state bodies in South Korea concerned with national prestige, funding agencies accountable to government masters, educational institutions bent on international competition, and journalists intent on good stories. They came together in a complex symbiosis to create a distorted image of scientific achievement.

Korea Post issued a postage stamp in Hwang's honour on 12 February 2005. Designed by Roh Jung-hwa, with a denomination of 220 won and printed in a quantity of 1.6 million, it was, the stamp tells us, specially issued to commemorate "the successful establishment of human cloned embryonic stem cells".

The previous stamp released by Korea Post

had shown nature on the island of Marado, with images of happy fish in azure seas, and the next was dedicated to the centenary of Rotary International. A further stamp for 2005 marked the sixtieth anniversary of Korean liberation from Japanese rule, an event of sufficient moment for any national postal service to celebrate. Later in the year, fusion was the subject — not fusion of the scientific kind, but of global cultures, symbolized by a knife and fork being handled as if they were chopsticks. Such was the heady philatelic setting of Hwang's stamp.

The long rectangle of the stamp contains on the left the expected graphics of high-tech cellular science, including the tip of a needle about to break through the wall of a human egg. A man in a wheelchair, silhouetted against the red background, rises triumphantly across the stamp to the right. He runs and leaps for joy in front of a purple cellular sun, before throwing himself into a woman's eager embrace. In a country with a large Christian population, the image of the cripple rising to walk carries clear connotations.

The tone of the scientific imagery on the left of the stamp is similar to that of James Brooke's article in *The New York Times* on

31 May 2005. "In the shadows of a darkened laboratory, a technician in a blue jumpsuit prodded and probed the egg's outer membrane...seeking to introduce a skin cell from a patient with an immune deficiency.

"Finally, on the third probe, the rubbery wall gave way. Magnified 250 times on a black-and-white screen, the egg could be seen making room for the new skin cell, with its new genetic code."

The right part of the stamp encapsulates the claims made in *Time*'s profile: "Hwang and his team at Seoul National University became the first to clone human embryos capable of yielding viable stem cells that might one day cure countless diseases." The stamp's implicit claims for a panacea for debilitating illnesses is just one of a vast number that make their way into the media when stem-cell research and human genomics are discussed.

The question raised by the stamp and other such visual and verbal hype is whether it is now possible to become a big beast in the international jungle of science without becoming ensnared in the perilous mechanisms of celebrity.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK.

A better than perfect match

Entanglement, a mind-boggling form of correlations that exist between objects in the quantum world, is helping to explain phenomena and jazzing up computing. But it looks as if much more may be in store.

Vlatko Vedral

The concept of correlations is familiar to us from everyday experiences. We try to eat healthy food, drink less coffee and give up smoking because we believe that such behaviour is correlated to a higher life expectancy. We send our children to good schools in the belief that one day this will correlate with them having a better quality of life. Ultimately, as is the case for any living entity, it is our ability to recognize and adapt to these correlations that maximizes our chances of survival.

But why are events in nature correlated in the first place? We don't really have an answer to this, although we are at least able to confirm the presence of correlations experimentally. Indeed, some of the most general and persistent correlations have been termed 'laws of nature', of which the second law of thermodynamics is a famous example. This states that things always flow from a more ordered to a less ordered state of affairs.

One would think we were fairly safe in assuming that correlations never exceed 100% (that is, that they cannot be better than perfect). If all children who go to good schools perform magnificently in life, we would say there is a 100% correlation between good schools and a successful life. You surely cannot get a better a correlation than this. But, odd as it may sound, correlations in nature can actually be better than perfect. This was first realized when physicists tried to infer the laws governing the behaviour of small objects; that is, in the study of quantum physics.

To qualify this paradox, imagine a simple two-state quantum system, such as the spin of an electron. Electrons are like small spinning-tops; each one rotates in its own way depending on the external circumstances. Just like a spinning-top, an electron can spin clockwise or counterclockwise in any given direction: horizontally, vertically, at 45°, and so on. Astonishingly, if we measure the electron spin at two different times, the correlations between these measurements can actually exceed any correlations allowed by classical physics.

Classically, spins at different times are expected to be correlated in the horizontal or vertical direction, so that if the first measurement of spin yields 'horizontal clockwise' so does the second. Real electrons, on

the other hand, behave quantum mechanically; their spin measurements can be correlated in the vertical direction at the same time as in the horizontal direction (and all other directions!). This is because electrons can spin simultaneously in the clockwise and counterclockwise direction — something that no spinning-top can do.

As a consequence of this, two electrons can be more correlated in spin than anything allowed by classical physics. This is not just a neat mathematical trick but an effect that has been confirmed experimen-



Electrons can be more correlated in spin than spinning-tops.

tally many times. Such quantum correlations that exist between objects and events are known as 'entanglement'.

A big question is whether entanglement is a wholly microscopic quirk, or whether it has any place in the large (macroscopic) world. Historically, this question has led to many apparent paradoxes in the quantum description of nature, but we now know that the answer is definitely 'yes'. For example, the magnetic behaviour of various solids is determined by the response of the electrons in the solid to the external magnetic field; the more correlated the spins of the electrons, the more they spin in the direction of the external field. A number of recorded measurements of magnetic responses of solids can only be explained if we assume that the electrons are entangled.

More surprisingly, this correlation between the spins of electrons and the direction of the external field happens in thermal equilibrium and at temperatures of up to 200 K, making these observations very real indeed. Equilibrium implies that the system is not being externally driven into an entangled state but is only subject to environmental noise. Moreover, systems at higher temperatures experience more noise, which should destroy correlations.

There is a sense in which entangled quantum objects can be thought of as a single quasi-object. This way of looking at things is paramount to understanding many phenomena such as Bose condensation (the formation of very low-temperature fluids), superfluidity and superconductivity.

Recently, we have also realized that quantum entanglement allows us to process information more efficiently than can be done with conventional (classical) computers. Certain tasks, such as the factorization of large numbers, can be performed exponentially faster on a quantum computer than on any classical computer. Whether such quantum correlations are already being used by living systems is not yet known, but there are some compelling indications that the answer might just again be 'yes'.

So, correlations are fundamental for our description and understanding of the world. In fact, it's tempting to say that things and events have no meaning in themselves, and that only the correlations between them are 'real'. This philosophy is known under the general name of 'relationalism'; Einstein's general theory of relativity is a shining example of its practical application in physics. Einstein was, however, unable to take relationalism one step further — to find a unified framework for general relativity and quantum physics. To this day this remains an open problem.

Could it be that this unification is proving elusive because quantum correlations (entanglement) have not been incorporated at the most fundamental level of its description? A view is now emerging according to which points in space and time can be thought of as correlated quantum objects, just like electrons in a typical solid. Can spatial and temporal distances between these points then be described as the amount of entanglement between them? Whatever the answer, one thing is certain: when it comes to entanglement, we have only just uncovered the tip of the iceberg.

Vlatko Vedral is professor of quantum information science at the University of Leeds, UK.

FURTHER READING

Taylor, J. C. *Hidden Unity in Nature's Laws* (Cambridge Univ. Press, 2001).
Born, M. *Natural Philosophy of Cause and Chance* (Dover, 1964).
Brooks, M. *New Sci.* **181**, 32 (2004).

COMSTOCK IMAGES/ALAMY

NEWS & VIEWS

PHYSIOLOGY

Plants on a different scale

Lars O. Hedin

Is there a unified theory that relates size and metabolic rate across all organisms? Maybe not, according to the results of experiments that measured respiration in plants of widely varying mass.

It is not every day that a 'law of nature' is challenged by empirical data, but on page 457 of this issue¹ we find just that. Peter Reich and colleagues¹ offer convincing evidence that refutes the idea that Kleiber's law — the all-encompassing prediction that metabolic rate should scale as the $3/4$ power of size across animals — can be extended to vascular plants (Fig. 1). Their findings question our understanding of how plant metabolism is organized. More broadly, they raise serious doubts about the notion^{2–4} that there exists a unified metabolic theory of scaling based on mechanisms that apply equally across plants and animals.

Here is the basic plot. In 1932, Max Kleiber showed that metabolic rates of mammals and birds scale as the $3/4$ power of body mass⁵. Kleiber's law has since become a standard of biology textbooks, demonstrating that mammals as different as mice, men and elephants obey the same quantitative metabolic relationship: a single line with a $3/4$ slope on a logarithmic plot of respiration versus body mass (Fig. 2a, overleaf). In 1960, Axel Hemmingsen⁶ showed that this $3/4$ slope applies even across a variation of six orders of magnitude in body size of mammals, fish, reptiles, insects and selected unicellular organisms. This extraordinarily broad metabolic pattern has all the makings of a universal biological principle⁷.

It has proven more difficult to identify the biophysical mechanisms responsible for the observed $3/4$ scaling slope. But Geoffrey West, James Brown and Brian Enquist have proposed a daring theoretical model^{2–4} to explain the $3/4$ metabolic exponent specifically, and the apparent ubiquity of quarter-power scaling relationships in biology generally. At its heart, this "metabolic theory of ecology"⁴ assumes that metabolism is constrained by resource delivery through internal branching networks in organisms. In the case of mammals, this means a circulatory system that supplies oxygen and nutrients to tissues. In the case of vascular plants, it means a system that delivers water and nutrients throughout the plant. If one assumes that natural selection acts to minimize transport costs through such networks, and given a set of rules for fractal branching,



R. MORSCH/CORBIS

Figure 1 | Little and large in the plant world. Reich *et al.*¹ find that Kleiber's law relating metabolic rate and mass does not apply to vascular plants across the six orders of magnitude in mass that they tested. One outstanding question, however, is whether their results extend to mature trees.

the model predicts that $3/4$ metabolic scaling ought to apply universally to all organisms that depend on such transport networks.

But here the plot thickens. Questions have emerged about whether the model is mathematically and conceptually consistent^{8,9}, and whether the best-fit regression slope is $3/4$, or $2/3$, or even variable across taxonomic groups¹⁰. (The idea¹¹ of $2/3$ metabolic scaling based on surface-to-volume considerations is an old one, predating Kleiber's monograph⁵ by five decades.) This current debate is largely focused on animals, for which there is an extensive data set of metabolism across species of varying sizes.

In contrast, direct measures of whole-plant respiration have been surprisingly rare. This means that early confirmations of the $3/4$ scaling prediction for plants were based either on indirect evidence³, which assumed that whole-plant transport of xylem water could substitute for metabolism (an assumption that has been questioned elsewhere^{12,13}), or on few samples, few species and few individuals¹⁴.

This is where Reich and colleagues¹ enter the picture. They report direct measurements

of whole-plant respiration, measured as carbon dioxide efflux in the dark (so ruling out the confounding factor of light-driven photosynthesis). Their results cover some 500 individual plants, across 43 species, from varying environments, and across six orders of magnitude variation in plant mass. The extent and quality of this new data set is refreshing, as statistical estimates of scaling slopes are notoriously sensitive to small sample sizes, or to heterogeneity in sampling techniques.

Reich and colleagues' results are astounding. Whole-plant respiration failed to fit the $3/4$ slope predicted by metabolic scaling theory. Rather, respiration scaled isometrically (that is, with a slope of 1.0 on a logarithmic plot) against plant mass. Mathematically, this means that, on a linear scale, respiration changes in direct proportion to variations in plant mass, as opposed to the more complex logarithmic relationship defined by the $3/4$ power function of Kleiber's law (Fig. 2b). Such a proportionate response in respiration against mass is perhaps intellectually less interesting, as it does not demand the kind of physiological complexity implied by $3/4$ scaling and metabolic scaling theory^{2–4}. Reich

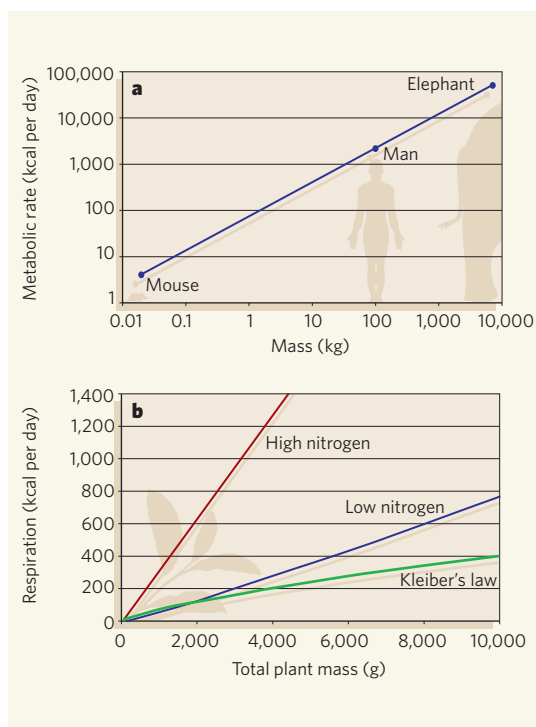


Figure 2 | Metabolism and mass.

a, Kleiber's law. Logarithmically, the basal metabolic rate of mammals varies with body mass as a straight line with a slope of 0.74. **b**, When expressed on a linear scale, the 3/4 scaling relationship of Kleiber's law (green line) shows total respiration rate decreasing proportionately as a function of body mass. In contrast, the isometric (slope 1.0) logarithmic scaling relationships described by Reich *et al.*¹ for vascular plants are consistent with a strictly proportional (linear) relationship between whole-plant respiration and mass. The authors also report that the intercept of the log-log scaling relationship depends on plant nitrogen. On the linear scale shown here, this translates to steep proportionality for plants in nutrient-rich greenhouse soils (red line) and less-steep proportionality for plants in nutrient-poor natural soils (blue line). Note that it is difficult to distinguish visually between isometric and 3/4 scaling lines on a linear scale, as this projection draws attention to only a limited part of the dynamic range.

et al. also report that the intercept of the scaling line differed between plants grown in the greenhouse and in the field, showing that respiration per unit mass differed systematically between these two environments.

The next observation is even more intriguing. When Reich *et al.* expressed plant respiration against the whole-plant content of nitrogen (instead of plant mass), the dissimilarity in intercept between environments entirely disappeared, but the overall scaling slope remained isometric. This suggests that the systematic difference in respiration per unit mass was caused by higher supplies of nitrogen in greenhouse compared with field environments. More importantly, across the different environments plant respiration was consistently more closely linked to variations in nitrogen than in mass. This result challenges not only the idea of a universal 3/4 scaling law, but also the notion that size alone is the dominant determinant of differences in metabolism across species. In hindsight, this need to explicitly consider nutrients might not be so surprising given that nitrogen is an essential component of enzymes such as Rubisco (responsible for carbon fixation in plants), and of proteins, chlorophyll and other biomolecules involved in plant photosynthesis and respiration.

Do the results simply imply that it is the business ends of plants — leaves and roots — in which most respiration occurs, and therefore most nitrogen is stored? In such a case, the metabolism versus nitrogen relationship might largely reflect rather pedestrian proportionate variations in crown and root volumes across individuals and species. A key issue is whether Reich and colleagues' findings extend beyond the size of tree saplings, the largest individuals sampled. Nitrogen allocation may

very well change as trees mature within natural forests and become increasingly subject to constraints of self-thinning¹⁵ and competition for nutrients¹⁶. A second question is why nitrogen emerges as such a strong correlate of metabolism, when phosphorus is generally considered the better predictor¹⁷.

Nonetheless, we find ourselves with a theory challenged. There is no question of the value

of ideas such as those of West and colleagues^{2–4}, as they dare to invoke first-principle universal mechanisms. But these new findings question a central tenet of the metabolic scaling theory — that size-dependent distribution networks exert the primary constraint on metabolic rates in vascular plants. Reich and colleagues' results¹ will spark considerable debate among ecologists and physiologists: at stake is the issue of whether there is a truly unified theory of metabolism that encompasses all organisms. ■ Lars O. Hedin is in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08540, USA. e-mail: lhedin@princeton.edu

1. Reich, P. B., Tjoelker, M. G., Machado, J.-L. & Oleksyn, J. *Nature* **439**, 457–461 (2006).
2. West, G. B., Brown, J. H. & Enquist, B. J. *Science* **276**, 122–126 (1997).
3. Enquist, B. J., Brown, J. H. & West, G. B. *Nature* **395**, 163–165 (1998).
4. Brown, J. H. *et al.* *Ecology* **85**, 1771–1789 (2004).
5. Kleiber, M. *Hilgardia* **6**, 315–353 (1932).
6. Hemmingsson, A. M. *Rep. Steno Meml Hosp. Nordisk Inst. Lab.* **9**, 6–110 (1960).
7. Smil, V. *Nature* **403**, 597 (2000).
8. Kozłowski, J. & Konarzewski, M. *Funct. Ecol.* **18**, 283–289 (2004).
9. Cyr, H. & Walker, S. C. *Ecology* **85**, 1802–1804 (2004).
10. White, C. R. & Seymour, R. S. *Proc. Natl Acad. Sci. USA* **100**, 4046–4049 (2003).
11. Rubner, M. Z. *Biol.* **19**, 536–562 (1883).
12. Midgley, J. J. *Trends Ecol. Evol.* **18**, 5–6 (2003).
13. Li, H. T., Han, X. G. & Wu, J. G. *J. Integr. Plant Biol.* **47**, 1173–1183 (2005).
14. Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M. & Charnov, E. L. *Science* **293**, 2248–2251 (2001).
15. Yoda, K., Kiara, T., Igawa, H. & Hozumi, K. *J. Biol. Osaka City Univ.* **14**, 107–129 (2003).
16. Chapin, F. S. *Annu. Rev. Ecol. Syst.* **11**, 233–260 (1980).
17. Elser, J. J. *et al.* *Ecol. Lett.* **6**, 936–943 (2003).

EXTRASOLAR PLANETS

Light through a gravitational lens

Didier Queloz

A planet with a mass lower than that of Neptune has been detected as its gravity bent the light from a remote star. This lensing technique adds to our arsenal for spotting small planets outside the Solar System.

Ten years ago, Michel Mayor and I discovered the first planet outside the Solar System orbiting a Sun-like star¹. This 'exoplanet' orbits the star 51 Pegasi in 4 days and has about the same mass as Jupiter (more than 300 times that of Earth). Its existence was revealed through highly accurate measurements of a tiny 'Doppler' variation in 51 Pegasi's radial velocity — the speed at which its position changes relative to an observer. This variation is caused by the gravitational pull of an orbiting planet with a mass 1,000 times less than that of 51 Pegasi. The discovery triggered many more Doppler surveys to search for planets around nearby stars. As a result, more than 160 planets with masses ranging from ten times that of Jupiter down to that of Neptune (which

is around 17 times Earth's mass) have been found (Fig. 1).

The range of planet masses and orbital parameters that can be identified by the Doppler technique is limited both by the detection sensitivity of the technique and by the time needed to survey at least one orbital period of each planet. The largest Doppler variation is caused by planets of high mass and those on short orbits. Therefore, low-mass planets can be detected only if they are on short orbits.

On page 437 of this issue, Beaulieu *et al.*² report the use of a different technique, known as gravitational microlensing, to detect a planet with a mass less than that of Neptune. Its orbital distance is about 2 AU (1 AU, or astronomical unit, is the Earth–Sun distance),

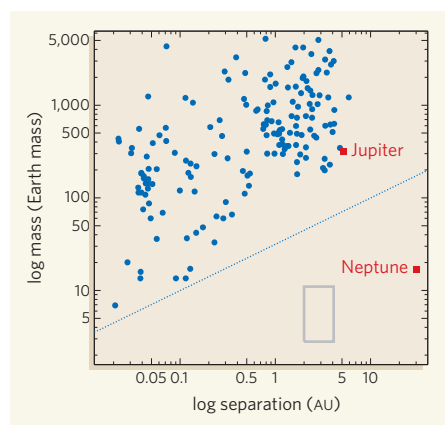


Figure 1 | Limits of the Doppler technique. Blue dots indicate the masses of the extrasolar planets found so far using the Doppler technique, plotted against their separation from their star. High-mass planets on short orbits (and so at a smaller separation) cause the greatest Doppler variation of the star's radial velocity. There is a detection threshold of accurate Doppler surveys of around 3 m s^{-1} radial velocity (blue line) below which no planet can be found using this technique. The grey rectangle indicates the region (including uncertainties) of the planet found by Beaulieu *et al.*² using the gravitational microlensing technique. The new planet is thus well below the Doppler sensitivity threshold.

which would correspond to an orbit between those of Mars and Jupiter in the Solar System. The planet's orbit is thus considerably larger than those found for planets of similarly small mass using the Doppler technique, for which the orbital distances are no more than 0.15 AU.

The evidence for Beaulieu and colleagues' planet lies in a small amplification effect that occurs when light from a remote, background star is deflected, as in the action of a lens, by the gravity of a closer, moving object as this approaches our line of sight to the star (Fig. 2). The brightness of the background star increases continuously and predictably over a period, generally a few weeks, and then decreases as the lensing object moves away again. This effect is generally caused by an intervening star, but sometimes a companion planet to this lensing star can also produce a detectable signal. In looking for such a signal, the rise in brightness corresponding to the approach of the star alerts observers to the need for more intensive observations. The aim is to detect a smaller brightness change that occurs on a timescale of around a day: the signature of an orbiting planet around the lensing star.

This planetary microlensing signal is challenging to detect. If seen and carefully traced,

however, it provides an unambiguous measurement of the mass ratio between the planet and the star, as well as supplying a measure of the separation of the planet from its star that depends on both the distance of the lensing star from Earth and its transverse velocity. The real mass and separation can then be obtained through a statistical approach. This uses a galactic model to estimate the expected distances and velocities of both the lens and the source star on the basis of the values anticipated for stars of their brightness in that sector of the sky.

From the wealth of planets found so far, we know that the frequency with which they are encountered increases with decreasing mass³. But Doppler detection of planets with masses lower than that of Neptune is still a challenge, and is restricted to planets on orbits shorter than a couple of months. The Neptune-mass planets with short-period orbits found by Doppler surveys^{4–7} suggest a relatively high occurrence of such low-mass planets orbiting normal stars. Beaulieu and colleagues' detection² further strengthens this suggestion: they argue that the probability of detecting a microlensing event from a planet such as theirs is more than ten times smaller than that of detecting an event caused by a giant planet.

Interestingly, the physical structure and formation history of this new planet, because of its orbital distance, is likely to be different from that of the closer low-mass planets discovered by the Doppler surveys. Whereas these short-period, Neptune-mass planets are likely to have migrated inwards and are now enduring the strong irradiation⁸ of their host star, causing some evaporation⁸ of their atmosphere, Beaulieu and colleagues' planet is certainly more akin to the ice giants Uranus and Neptune of our Solar System.

The discovery by Beaulieu *et al.*² demonstrates that microlensing surveys can detect sub-Neptune planets with low separations from their stars. In combination with other planet-finding techniques, microlensing surveys will play their part in eventually gathering a complete picture of planet statistics right down to the regime of rocky planets. A deeper insight into the physical models of planetary formation for low-mass giant and high-mass rocky planets will surely follow.

Didier Queloz is at the Observatoire de Genève, 51 chemin des Maillettes, 1290 Sauverny, Switzerland.
e-mail: didier.queloz@obs.unige.ch

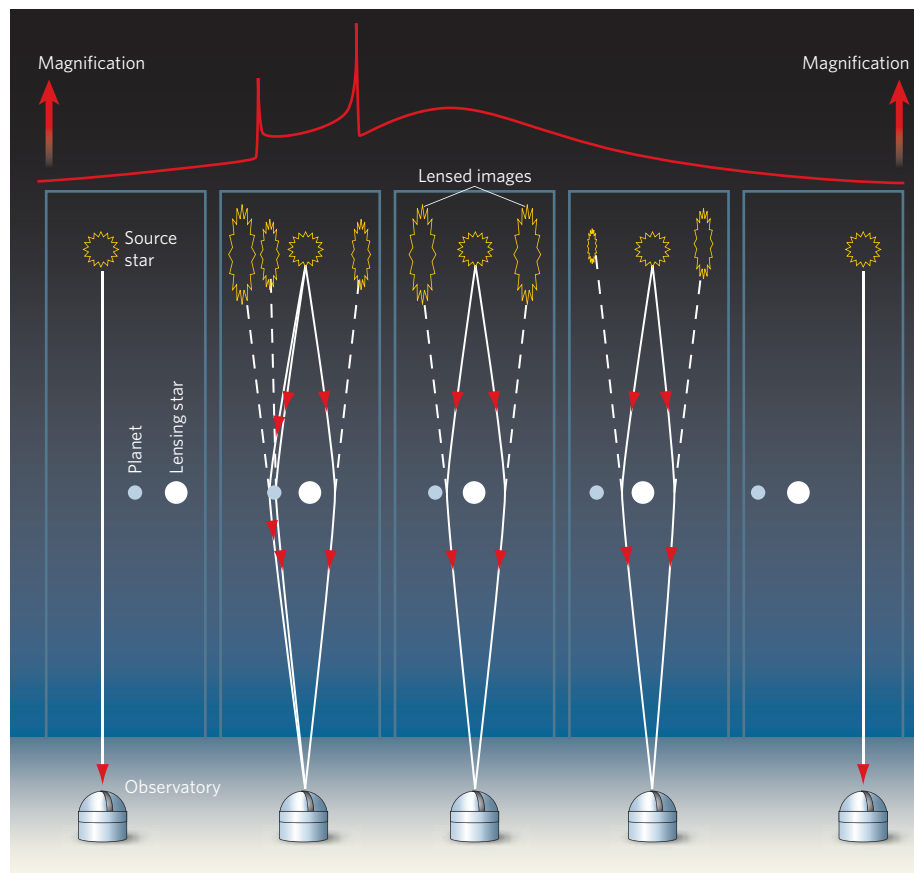


Figure 2 | Lensing planet. The gravitational microlensing effect exploited by Beaulieu *et al.*² results from the bending of space-time near an object of given mass that is predicted by Einstein's general theory of relativity. An object, such as a star, crossing our line of sight to a more distant source star will affect the light from that star just like a lens, producing two close images whose total brightness is enhanced. If the lensing star is accompanied by a planet, one can (potentially) observe not only the principal effect from the star, but also a secondary, smaller effect resulting from perturbation by the planet.

- Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
- Beaulieu, J.-P. *et al.* *Nature* **439**, 437–440 (2006).
- Santos, N. *et al.* *ESO Messenger* **110**, 32 (2002).
- Rivera, E. J. *et al.* *Astrophys. J.* **634**, 625–640 (2005).
- Santos, N. C. *et al.* *Astron. Astrophys.* **426**, L19–L23 (2004).
- McArthur, B. E. *et al.* *Astrophys. J.* **614**, L81–L84 (2004).
- Bonfils, X. *et al.* *Astron. Astrophys.* **443**, L15–L36 (2005).
- Baraffe, I. *et al.* *Astron. Astrophys.* **436**, L47–L52 (2005).

METABOLISM

Bile acids heat things up

John D. Baxter and Paul Webb

Thyroid hormone causes fat loss, but harnessing this action to treat obesity is difficult because it is associated with harmful side effects. However, bile acids generate active thyroid hormone just where it is needed.

Obesity is pandemic in the Western world and is becoming a problem in other countries¹. Stroke, heart attack, diabetes, kidney failure, arthritis and peripheral vascular disease are all consequences of this disorder. But the outlook is not all bleak. If obesity can be reduced, there should be enormous health benefits — even a 5% decrease in body weight through diet and exercise cuts the risk of type 2 diabetes by more than 50%². Nevertheless, efforts to address the problem by encouraging lifestyle modifications have met with limited success. So major efforts are under way to find other therapies. In this issue, Watanabe *et al.* (page 484)³ report a previously unappreciated role for bile acids in regulating fat that might also be exploited pharmacologically to tackle obesity.

Body-weight control is essentially a question of energy balance: if we eat more calories than we expend, we become fat⁴. Many therapies, therefore, aim to reduce food intake. Surgery has had impressive results, but carries substantial risks. Current pharmacological manipulation of the pathways that regulate appetite or block dietary fat absorption have

limited utility. High-fat, fast-food diets can override delicate regulatory networks established by appetite-regulating hormones, and it has been difficult to develop drugs that reduce food intake in the face of powerful environmental influences. Moreover, drugs that block fat absorption have limited efficacy.

Are there other approaches to combat obesity? Endocrinologists have long known that hormones can alter the body's fat content. And although they cannot override the laws of thermodynamics, hormones do alter the energy-balance equation by increasing metabolic rate. An illustration of this principle is seen in disorders involving excess thyroid hormone, where stimulation of metabolic rate and decrease in body fat can occur even with increased dietary intake. Thyroid hormone itself cannot be used to treat obesity because of deleterious side effects that can include increased heart rate, atrial arrhythmias and breakdown of muscle and bone. But manipulating this pathway may be useful, and selective analogues of thyroid hormone are already showing promise in reducing body fat in animals⁵.

Watanabe *et al.*³ now report that bile acids have a role in regulating thyroid hormone signalling and energy homeostasis. Although the classical role for bile acids is to enhance fat absorption in the intestine, they also regulate the metabolism of fat by acting as signalling molecules through pathways involving the cell-surface receptor TGR5 (ref. 6) and nuclear receptors such as FXR (ref. 7). Watanabe *et al.* show that bile acids, which include cholic acid, stimulate production of active thyroid hormone in fat cells (Fig. 1).

Bile acids have been shown to mitigate diet-induced obesity, but this work⁸ has received limited attention. Watanabe *et al.* confirm and expand this observation, showing that feeding mice with cholic acid moderates the effects of a high-fat diet. The animals were less obese and their blood glucose levels were better regulated than in controls fed the high-fat diet alone. Cholic acid had no effect on body weight in lean animals, and mice fed cholic acid did not eat less. Instead, bile acids reduced weight gain by increasing energy expenditure: the treated mice consumed more oxygen and produced more carbon dioxide, and so had a decreased respiratory quotient, indicative of increased fat burning.

The basis for this remarkable effect became clear through genomics. Large-scale gene-expression studies showed that bile acids increase the expression of several genes involved in energy metabolism in brown fat tissue but not in liver. One of these genes encodes the type 2 deiodinase enzyme (D2), which converts the minimally active precursor of thyroid hormone, thyroxine (T₄), into the

TRAVEL

Fitting the bill

Anecdotal evidence about human travel is plentiful. But quantifying human movement and dispersal, and then applying general principles to those data, is not straightforward. Elsewhere in this issue, a group of theoretical physicists describe how they have taken an ingenious approach to the problem (D. Brockmann *et al.* *Nature* **439**, 462–465; 2006). They use the dispersal of dollar bills within the United States as a proxy measurement for human movement.

The researchers analysed data on the peregrinations of more than half-a-million US dollar bills recorded over a five-year period on an online bill-tracking system (www.wheresgeorge.com). This allowed them to quantify even statistically rare events with high reliability. They found that the movement of dollar bills resembles

enhanced diffusion, or 'superdiffusion'. Long-distance jumps are disproportionately important in the distribution of travelling distances. And this distribution decays as a power law reminiscent of 'Lévy flights', which are characterized by many short steps interspersed with long-distance jumps.

The result may seem unsurprising — most cash transactions are carried out locally, but every now and again someone departs and uses the bill at a distant location. The beauty is that actual numbers can now be put on the process.

Although the observed distribution of travelling distances implies superdiffusion, however, this process is attenuated by the tendency of bills to remain in the same area for longer than might be expected given the



PARAMOUNT/THE KOBAL COLLECTION

overall pattern of movement. Thus, human travelling behaviour apparently involves disproportionately long waiting times between displacements as well as jumps without any characteristic distance scale.

But how well do bill trajectories reflect actual human movement? Happily, the authors find that data on passenger travel on the US aviation network, and long-distance

human travel information published by the US Bureau of Transportation Statistics, agree well with the results based on dollar-bill trajectories.

This research is more than a curiosity — epidemiologists could in principle use quantitative information on human movement to understand better how infectious diseases such as influenza spread.

Rory Howlett

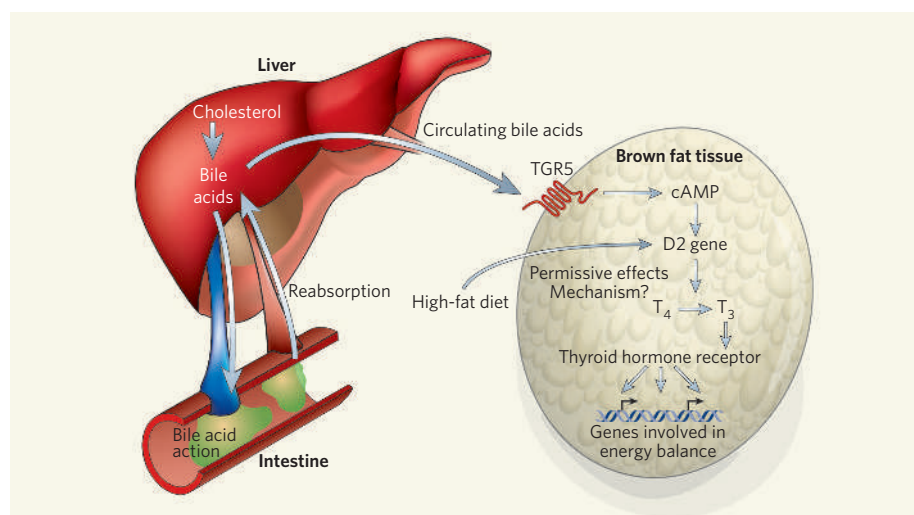


Figure 1 | Bile acids in energy homeostasis. Bile acids are synthesized from cholesterol in the liver, stored in the gallbladder, and secreted after meals to promote absorption of fat from the intestine. They are then either excreted or reabsorbed into the circulation. Watanabe *et al.*³ demonstrate that bile acids increase the metabolic rate in fat cells by binding to a G-coupled protein receptor (TGR5) that increases cAMP content and induces D2 expression, thereby enhancing local conversion of T_4 to the active T_3 . These effects are observed only in animals that are fed a high-fat diet, as this sensitizes the D2 response to bile acids through an unknown mechanism.

potent form, triiodothyronine (T_3). Direct evidence that D2 (and therefore T_3) is involved in the actions of cholic acid came from mice in which the D2 gene had been disrupted, and which showed none of the effects seen in animals fed cholic acid. Increased metabolic rate in the brown fat of rodents through D2-induced increases in T_3 is crucial for sustaining body temperature in cold conditions (adaptive thermogenesis). So bile acids seem to block fat accumulation by inducing thyroid hormone signalling in brown fat.

Watanabe *et al.* have begun to tease apart the mechanism behind the effects of D2 on cholic acid. They show that bile acids regulate D2 expression by binding to TGR5, which is present in high levels in brown fat, and rule out the possibility that FXR is involved. TGR5 belongs to a family of receptors known as G-protein-coupled receptors. It increases intracellular cyclic AMP (cAMP), which is already known to increase D2 levels. The authors may also have explained why the bile-acid effect is seen in obese and not lean animals: the induction of D2 in the brown fat from mice on the high-fat diet was more sensitive to activators of the cAMP pathway than in controls. However, the origin of this sensitivity is unclear.

Is this discovery relevant for humans? The tissues involved in thermogenesis are different in humans and rodents: adult humans have very little brown fat, but skeletal muscle is crucial for energy homeostasis. Human skeletal muscle expresses significant levels of D2, and Watanabe *et al.* detected TGR5 in cultured human skeletal muscle cells. Treatment with bile acid increased metabolic rate in these cultures. The authors note that a human variant of the D2 gene is associated with a decreased rate of whole-body glucose disposal, a rate

that is mostly determined by glucose uptake into muscle. So there are hints that D2 regulates metabolism in skeletal muscle in humans in a similar manner to that in brown fat in mice.

The bile-acid pathway might be important from both physiological and pharmacological perspectives. Postprandial concentrations of bile acids should be sufficient to stimulate cAMP production and D2. So bile acids could be hormonal signals linking food intake to diet-induced increases in metabolic rate. Bile acids caused toxic side effects, especially in the liver, when they were tested in other disorders⁹, so any similar compounds designed for clinical use would need get around these problems. However, the finding that bile acids increase T_3 levels selectively in tissues where T_3 can promote energy expenditure provides a significant advance in our understanding of the regulation of energy homeostasis and a potential approach to fighting obesity.

John D. Baxter and Paul Webb are at the Diabetes Center and Department of Medicine, University of California School of Medicine, S1212, 513 Parnassus Avenue, San Francisco, California 94143-0540, USA.
e-mails: jbxter918@aol.com;
pwebb@diabetes.ucsf.edu

- Hedley, A. A. *et al.* *J. Am. Med. Assoc.* **291**, 2847–2850 (2004).
- Gaede, P. *et al.* *N. Engl. J. Med.* **348**, 383–393 (2003).
- Watanabe, M. *et al.* *Nature* **439**, 484–489 (2006).
- Spiegelman, B. M. & Flier, J. S. *Cell* **104**, 531–544 (2001).
- Baxter, J. D. *et al.* *Trends Endocrinol. Metab.* **15**, 154–157 (2004).
- Kawamata, Y. *et al.* *J. Biol. Chem.* **278**, 9435–9440 (2003).
- Kalaany, N. Y. & Mangelsdorf, D. J. *Annu. Rev. Physiol.* doi:10.1146/annurev.physiol.68.033104.152158 (2005).
- Ikemoto, S. *et al.* *Am. J. Physiol.* **273**, 37–45 (1997).
- Shoenfeld, L. J. & Lachin, J. M. *Ann. Intern. Med.* **95**, 257–282 (1981).



50 YEARS AGO

Among the strangest forms of animal behaviour is that of the honey-guides, African birds distantly related to the American woodpeckers, which 'guide' men, baboons and rats (honey-badgers) to the nests of wild honeybees—supposedly so that these nests can be broken open. A study of the behaviour is described by Dr. Herbert Friedmann, U.S. National Museum curator of birds, in a bulletin issued by the Smithsonian Institution, Washington... "When the bird is ready to begin guiding, it either comes to the person and starts a repetitive series of churring notes or it stays where it begins calling these notes and waits for the human to approach it more closely... As the person comes to within 15 to 50 feet from it, the bird flies off with an initial conspicuous downward dip, and then goes off to another tree, not necessarily in sight of the follower; it is more often out of sight than not. It waits there, churring loudly until the follower again nears it, when the action is repeated. This goes on until the vicinity of the bees' nest is reached... It waits for the follower to open the hive and usually remains until the person has departed with a collection of honeycomb, when it comes down to the plundered nest and begins to feed on the bits of comb left strewn about."

From *Nature* 28 January 1956.

100 YEARS AGO

"Sounding Stones" — It may be of interest to add to the list of musical stones provided by your correspondents another limestone, viz. the very hard, crystallized, coral rock of the coasts of British East Africa. Among the bizarre forms assumed by these rocks under the erosion of the sea, isolated pillars with projecting arm at the top, like a gallows or an inverted capital "L," are common in places. This horizontal arm in many cases gives a clear musical note when struck with a stone or hammer, being thus a ready suspended natural gong.

From *Nature* 25 January 1906.

50 & 100 YEARS AGO

DEVELOPMENT

Twists of fate in the brain

Richard V. Pearce II and Clifford J. Tabin

How does the complex array of cell types and functions in the mammalian brain develop? Tracking cells by gene expression shows how their fates derive from organization within the simple embryonic neural tube.

While forming functional domains in the developing brain, cells crawl considerable distances to reach their final destinations. So it has been extremely difficult to determine when and where during this process the neuronal cells 'learn' which distinct brain region they are fated to become. But advances in mouse genetic manipulation now allow researchers to indelibly mark a cell population at a discrete place and time in development, based on its gene-expression pattern, and to follow the cells' subsequent progress^{1,2}. A series of papers^{3–5} have exploited these innovations to trace the 'decisions' of cells to form certain brain centres (nuclei) back to instructions in the developing neural tube in the embryo.

The embryonic neural tube gives rise to the brain and spinal cord, and is one of the best-studied models of how spatial organization of a structure emerges during development. In the neural tube that will give rise to the spinal cord, fate specification takes place essentially along a two-dimensional cartesian grid. Distinct sets of transcription factors are expressed at various locations within this grid, establishing unique progenitor populations.

These factors regulate groups of target genes, setting in motion a developmental programme that defines the ultimate fate of cells derived from each progenitor population⁶. The progenitor domains can be repeatedly divided into ever smaller sub-domains, each defined by one or more transcription factor(s). The cells derived from these pools can either stay put or move short distances radially, while basically maintaining their dorsal–ventral relationships.

In the developing brain, the situation is much more complex. In many regions there are so-called germinal zones, which serve as a source of progenitor cells. These cells migrate along complicated pathways, both laterally and radially, ending up next to neighbours that are quite distinct from the cells that were near them in the germinal zone. Moreover, single progenitor cells within a germinal zone can give rise to daughter cells that end up in different regions and adopt dissimilar fates⁷.

The germinal zone known as the rhombic lip consists of the dorsal edge of the part of the neural tube that lies between the developing midbrain and what will become the spine. In this location the neural tube is open at the top although it has fused in rostrally (towards the nose) and caudally (towards the tail). As a

consequence, the edges of the neural tube bulge outward forming a rhombic fold on the top (Fig. 1a). This region begets cells that migrate extensively. For example, cells originating in the caudal rhombic lip contribute both to the pre-cerebellar nuclei, which extend neural connections to specialized neurons called granule cells, and to the inferior olive nucleus, which connects to the cerebellar Purkinje cells (neurons that transmit signals from the cerebellum to the rest of the brain). When and where the specification of individual cell types and their final locations occurs during these migratory processes has been completely unclear.

The recent papers, published in *Neuron*^{3–5}, build upon previous work^{8,9} to refine dramatically our map of the fates of rhombic-lip cells. Machold *et al.*³ demonstrate that precursor cells expressing the transcription factor Math1 have different fates depending on when they migrate from the rhombic lip (Fig. 1a, c). Those that migrate early from the rostral rhombic lip give rise to specific hindbrain and deep cerebellar nuclei, whereas those that move out later become cerebellar granule cells. The latter end up spreading progressively from the rostral to the caudal cerebellum as time goes on.

Wang *et al.*⁴ define four streams of migratory cells derived from Math1-expressing

progenitors. These emanate from the rostral and caudal rhombic lip and then branch to produce multiple nuclei in the brainstem and cerebellum. Both Wang *et al.* and Machold *et al.* demonstrate that the activity of Math1 itself is required for the development of many of these nuclei. So, the spatial and temporal expression of Math1 instructs the progenitor cells residing in the rhombic lip to become specific sets of brainstem and deep cerebellar nuclei.

A fuller understanding of the molecules that specify neural fates comes from Landsberg *et al.*⁵, who demonstrate that the rhombic lip is divided into discrete molecularly defined pools of progenitor cells along the dorsal–ventral axis. They show that the rhombic lip is marked out by a dorsal–ventral concentration gradient of the Wnt1 protein; the region is further subdivided into a series of smaller domains by the expression of distinct transcription factors (Fig. 1b).

Dorsally, the presence of the secreted molecule Gdf7 and the transcription factor Lmx1a define a set of progenitor cells fated to become the epithelia of the choroid plexus^{5,8}. Just ventral to this is the Math1 domain where cells will become distinct nuclei in the brainstem. Finally, and most ventrally, there is a sub-domain that contains Ngn1, which appears to mark out the cells that will go on to form the inferior olive nucleus^{5,9}. Additionally, the ratio of these precursor pools within the rhombic lip is carefully modulated by the transcription factor Pax6. Loss of Pax6 results in an increase in the numbers of the Ngn1-positive precursors at the expense of the Math1-positive precursors⁵.

Together, these papers vastly improve the molecular resolution of the rhombic-lip fate map, thereby providing important insights into how the complex structures of the brain

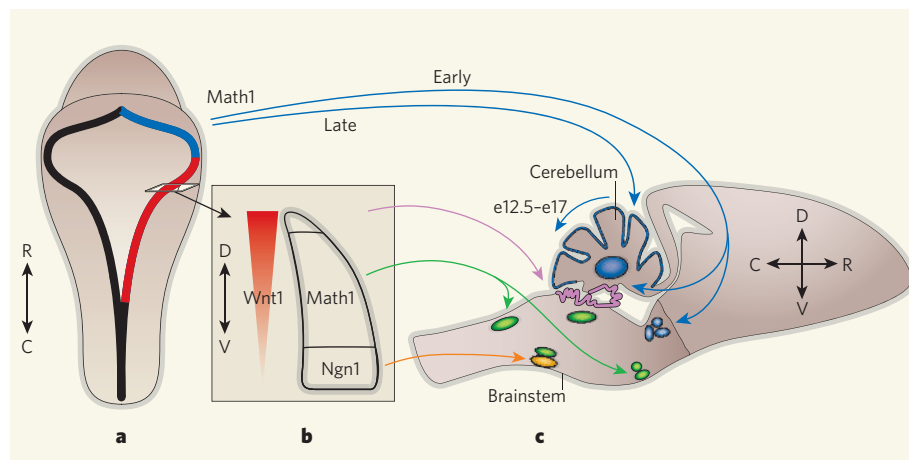


Figure 1 | Fate in the rhombic lip. **a**, The rhombic lip of the embryonic day 12.5 (e12.5) embryo, seen from the dorsal side, is divided into two domains: rostral (blue, nearest what will become the nose) and caudal (red, nearest what will become the back of the head). **b**, A dorsal–ventral cross-section through the e12.5 caudal lip shows the molecular domains that will contribute to distinct brainstem nuclei. **c**, The mature brain. Arrows show the migratory populations stemming from the sub-domains in the lip, and contributing to mature nuclei and granule cells. The inferior olive nucleus is highlighted in orange on the ventral side of the brainstem, and the choroid plexus is the folded purple structure linking the brainstem to the cerebellum. R, rostral; C, caudal; D, dorsal; V, ventral.

develop. They show that the fates of cells in the germinal zone of the rhombic lip are specified both spatially and temporally. Spatial specification comes from molecular domains distributed along the dorsal–ventral^{5,8,9} and rostral–caudal axes^{3,4,10} (such as those in the spinal neural tube), and temporal specification is established through sequential egress from those domains³ (Fig. 1).

The specification of progenitors before their dispersal from the germinal zone has a logic to it from a functional perspective. The collection of brain nuclei defined by these coordinated molecular activities, although spatially separated in the adult brain, together constitute components of the proprioceptive system. This is the sensory system that allows the positions of the limbs to be detected in the absence of visual cues. So this dispersed network of functionally cooperative brain structures can trace its origins back to the same continuous strip of germinal tissue.

These studies, particularly the work of Landsberg and colleagues⁵, unveil a precise, molecularly defined organization within the rhombic-lip germinal zone. This organization is in keeping with the model already established for the spinal neural tube, where neural sub-domains are characterized by the expression of specific transcription factors, established in response to graded intercellular signals, now extended to include the germinal zones of the brain.

The work relied on powerful genetic techniques for fate mapping developed in the past few years^{1,2}. Although accurate fate maps are crucial to our understanding of development, generating them has usually required unfettered access to the developing embryo, prohibiting *in vivo* mammalian fate mapping, and has been limited by the inability to molecularly define the cell population that was labelled initially. Also, because of the difficulty in garnering such data, the development of fate maps in mutant tissues has been unfeasible, so investigators could only speculate about how specific genetic modifications might alter the migratory routes and ultimate fate of specific cell types. As exemplified by the current studies, the new genetic approaches in mice allow cell fates to be correlated precisely with gene-expression domains in both normal and mutant embryos. It is clear that the ever-increasing pace of mammalian genetics will continue to refine our view of the molecules that regulate cellular fate throughout the embryo. ■

Richard V. Pearce II and Clifford J. Tabin are in the Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.
e-mail: tabin@genetics.med.harvard.edu

6. Jessell, T. M. *Nature Rev. Genet.* **1**, 20–29 (2000).
7. Lin, J. C. & Cepko, C. L. *J. Neurosci.* **21**, 159–168 (2001).
8. Currie, D. S., Cheng, X., Hsu, C. M. &

- Monuki, E. S. *Development* **132**, 3549–3559 (2005).
9. Hoshino, M. *et al. Neuron* **47**, 201–213 (2005).
10. Sgaier, S. K. *et al. Neuron* **45**, 27–40 (2005).

MATERIALS SCIENCE

A new order for metallic glasses

Alain Reza Yavari

Like normal glass, metallic glasses lack long-range order. But experiments and simulations show that, on the nanoscale, clusters of atoms interconnect in these materials to form highly structured ‘superclusters’.

Metallic glasses are peculiar metallic materials, usually alloys, that lack the long-range order of normal, crystalline metals. The attractive interactions, and differences in size, of atoms of different types in metallic glasses do, however, lead to a short-range order characterized by clusters of ‘solute’ atoms of one type surrounded by atoms of a more numerous species, the solvent. This much has been known for a long time. But just how these atomic clusters connect to fill space nearly as densely as crystalline solids with the same atomic composition has remained a mystery.

Some 45 years on from the discovery of the first metallic glasses¹, Sheng and colleagues (page 419 of this issue)² provide an essential missing piece of the puzzle, determining the three-dimensional structure of several metallic glasses that contain two different types of atoms. In these ‘binary’ glasses, they find nanoscale medium-range order, often consisting of closely packed icosahedral (20-faced) assemblies of some 13 neighbouring atomic clusters each centred on a solute atom.

The lack of long-range order in metallic glasses is signalled by the absence of sharp Bragg peaks — features characteristic of a periodically structured material — in the angular distribution of diffracted beams (X-rays, electrons and neutrons) used to probe them. These missing peaks led early researchers to consider atomic packing in metallic glasses to be similar to John Bernal’s ‘dense random packing’ of hard spheres in liquids³. In this picture, the larger solvent atoms are densely, randomly packed⁴, and the solute atoms are jammed into cavities left available to them by the geometry of this packing. Such models were later abandoned — in part because metallic glasses such as copper–zirconium (Cu₆₀Zr₄₀) were discovered whose solvent atoms were smaller than their solute atoms — in favour of packings of atom clusters in fixed ratios⁵ with short-range order similar to that of crystalline compounds of the same atomic composition. These ‘stereochemical’ models could reproduce the experimental diffraction patterns and the corresponding distributions (known as radial distribution functions) of nearest and next-to-nearest neighbour atomic

shells. But metallic glasses also exhibit order over 1–1.5 nanometres (for comparison, a typical metallic atom has a diameter of around 0.3 nm), and, until recently, no tangible description of how clusters of atoms interconnect to generate this medium-range order was available.

This situation changed with the proposal⁶ a couple of years ago of dense packings of overlapping atomic clusters as the fundamental scheme for metallic glasses. In this model, an arbitrary ‘face-centred-cubic’ (fcc) lattice of clusters is chosen to achieve high density. (The fcc lattice is the densest possible cubic packing, and consists of elements placed at each vertex, and in the middle of each face, of a cubic structure.) But to maintain long-range disorder — to keep the material glassy — a strain factor has to be introduced to limit the coherence of such a lattice to the 1–1.5-nm scale. This model has been successful in predicting the compositions of most glass-forming alloys, and alloys with lower melting points than any of their constituents, known as eutectics⁷.

Sheng *et al.*² set out to solve the three-dimensional structure of metallic glasses without resorting to a predetermined structural model. Instead, they use reverse Monte Carlo⁸ simulations based on experimental X-ray diffraction and absorption data, as well as *ab initio* simulations of molecular dynamics⁹, to determine the structure of a variety of binary nickel-based and zirconium-based metallic glasses with different ratios of atomic size and different levels of solute concentration. (Figure 1d on page 420 provides an example of the structure of a nickel–phosphorus glass, Ni₈₀P₂₀, ascertained through the reverse Monte Carlo method.) The radial distribution functions calculated using both simulation methods agree remarkably well with those measured directly using diffraction, and the mass (or atomic) densities they yield are within 1–2% of the experimentally measured values.

Knowing the three-dimensional positioning of the atoms of the glass allows the short- to medium-range details of its structure to be investigated. The number of nearest neighbours in an atomic cluster (the ‘coordination

1. Zinyk, D. L., Mercer, E. H., Harris, E., Anderson, D. J. & Joyner, A. L. *Curr. Biol.* **8**, 665–668 (1998).
2. Branda, C. S. & Dymecki, S. M. *Dev. Cell* **6**, 7–28 (2004).
3. Machold, R. & Fishell, G. *Neuron* **48**, 17–24 (2005).
4. Wang, Y. Y. *et al. Neuron* **48**, 31–43 (2005).
5. Landsberg, R. I. *et al. Neuron* **48**, 933–947 (2005).

number') can be determined using a technique known as Voronoi tessellation¹⁰, which involves the division of the glass's structure into regions centred on individual solute atoms. The basic units of the short-range order that emerges are various polyhedra¹¹ of around 9 to 13 atoms, with a solute atom in the middle. The precise form of these polyhedral clusters is controlled by the ratio of the effective sizes of solute and solvent atoms, and so changes according to the elemental constituents of the glass. Sheng and colleagues also find a moderate variation in coordination numbers (that is, a range of different, quasi-equivalent clusters) in the same material. This flexibility allows for a more efficient packing of 'soft' atoms without requiring an fcc structure.

Once the three-dimensional positions of the clusters have been mapped out, their topological packing can be determined by a technique known as common neighbour analysis¹². In three of the metallic glasses investigated by Sheng and colleagues² — $\text{Ni}_{80}\text{P}_{20}$, nickel-boron ($\text{Ni}_{81}\text{B}_{19}$) and zirconium-platinum ($\text{Zr}_{84}\text{Pt}_{16}$) — the clusters pack with appreciable icosahedral medium-range order, regardless of the short-range order within the clusters. Each solute-centred cluster shares its solvent atoms with about 12 neighbouring clusters, forming 'super-icosahedra' of 70–80 atoms that are about 1.5 nm wide, or fragments of such structures. As does the fcc-packing model⁶, this icosahedral packing of clusters generates cavities similar to those in the random dense packing model³, into which additional solute species of different sizes may be introduced¹³. This allows bulk metallic glasses to form^{14,15}.

Sheng and colleagues' *ab initio* calculations indicate that, as the solute content of the metallic glass is increased and solute-solute nearest neighbours become numerically unavoidable, the connection between solute atoms becomes string-like. When the solute strings percolate, as in nickel-niobium glass ($\text{Ni}_{63}\text{Nb}_{37}$), the solute-solute connection begins to resemble a network, and a spectrum of atomic packing schemes that varies with solute radius and concentration is generated.

The modelling techniques used by the authors do have some inherent limitations. *Ab initio* molecular dynamics simulations can still, for example, only be performed on limited time and length scales because of present limits on computational power. The short time-window also means that the cooling ('quench') rates currently used in simulations are more than 10^{12} K s^{-1} , far above laboratory quench rates of 10^3 – 10^6 K s^{-1} , restricting the ability of the atoms to sample all possible configurations as they cool.

Nevertheless, the work of Shen *et al.*² will serve to establish a firmer picture of the structure of metallic glasses. This information is fundamental to further applications involving these materials, which are already used commercially because of their exceptional magnetic and mechanical properties. More

extensive exploitation of their mechanical properties — in particular their elastic response up to 2% strain — depends on better understanding of their plastic deformation. This deformation, usually leading to failure of the material beyond the 2% elastic limit, is heterogeneous at ambient temperature¹⁶, and occurs in highly localized thin shear bands associated with tiny areas of collective atomic mobility ('shear transformation zones'¹⁷) that can percolate across the cross-section of the glass. Contrary to the hardening observed in crystalline materials under strain, heterogeneous deformation in metallic glasses has the opposite effect owing to the destruction of medium-range order. Sheng and colleagues' investigations on exactly this scale could thus provide further insight into the effects of deformation, and ways to improve the mechanical properties of metallic glasses. ■

Alain Reza Yavari is at the Laboratoire de Thermodynamique et Physico-Chimie Métallurgiques, Institut National Polytechnique de Grenoble, BP 75,

38402 St-Martin-d'Hères Campus, France, and at the Centre Nationale de la Recherche Scientifique (CNRS) and European Synchrotron Radiation Facilities, Grenoble, France.
e-mail: yavari@ltpcm.inpg.fr

1. Klement, W., Willens, R. H. & Duwez, P. *Nature* **187**, 869–871 (1960).
2. Sheng, H. W., Luo, W. K., Alamgir, F. M., Bai, J. M. & Ma, E. *Nature* **439**, 419–425 (2006).
3. Bernal, J. D. *Nature* **185**, 68–70 (1960).
4. Polk, D. E. *Acta Metall.* **20**, 485–489 (1972).
5. Gaskell, P. H. *Nature* **276**, 484–485 (1978).
6. Miracle, D. B. *Nature Mater.* **3**, 697–702 (2004).
7. Yavari, A. R. *Nature Mater.* **4**, 1–2 (2005).
8. McGreevy, R. L. *J. Phys. Condens. Matter* **13**, R877–R913 (2001).
9. Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A. & Joannopoulos, J. D. *Rev. Mod. Phys.* **64**, 1045–1097 (1992).
10. Finney, J. L. *Nature* **266**, 309–314 (1977).
11. Nelson, D. R. *Phys. Rev. B* **28**, 5515–5535 (1983).
12. Clarke, A. S. & Jónsson, H. *Phys. Rev. E* **47**, 3975–3984 (1993).
13. Greer, A. L. *Science* **267**, 1947–1953 (1995).
14. Inoue, A. *Acta Mater.* **48**, 279–285 (2000).
15. Johnson, W. L. *Mater. Res. Bull.* **24**, 42–56 (1999).
16. Spaepen, F. *Acta Metall.* **25**, 407–412 (1977).
17. Argon, A. *Acta Metall.* **27**, 47–53 (1979).

DNA REPAIR

Tails of histones lost

André Nussenzweig and Tanya Paull

A double-stranded break in DNA can profoundly destabilize a cell's genome. But how does the cell recognize the damage and halt division until it can be fixed? The answer lies in the proteins that package and unravel DNA.

DNA damage induces cell-cycle checkpoints that transiently arrest progression through the cell-division cycle. This delay gives the DNA-repair machinery sufficient time to fix genomic damage before the cell cycle resumes. Two studies, one by Tsukuda *et al.*¹ published at the end of last year and one by Keogh *et al.* in this issue (page 497)², demonstrate that modifications to the DNA packaging around the break site help to coordinate DNA repair with cell-cycle checkpoints.

Double-strand breaks in chromosomal DNA are repaired either through direct end-joining or through a process known as recombination in which the broken ends are spliced to the corresponding undamaged DNA on a sister chromosome, and the break is filled in using the undamaged DNA as a template. DNA breaks that are not rapidly rejoined are chewed back by exonuclease enzymes so that a length of single-stranded DNA hangs out from the remaining DNA. These single-stranded intermediates perform at least two crucial functions: the exposed single strands are bound by Rad51 proteins that initiate the search for a complementary template from which to repair them, and they act as a signal to arrest the cell cycle. Resumption of the cell cycle following checkpoint arrest is

generally concurrent with repair, suggesting that the elimination of single-stranded intermediates may also control exit from the checkpoint.

In the nucleus, DNA is wrapped about histone proteins to create nucleosomes, and then it is further twisted up into higher levels of packaging. This tight packing probably creates a structural barrier to molecules that recognize and respond to DNA damage. This problem seems to be solved by proteins that control the organization of the chromatin (that is, DNA and its associated proteins). Complexes of such chromatin-remodelling proteins have been found near double-strand breaks; for example, the INO80, SWR1 and NuA4 complexes in budding yeast bind within two kilobases of a break, and are needed for the DNA to be repaired^{3–5}. It is proposed that the chromatin-remodelling complexes might unravel the packaging, allowing repair enzymes access to the DNA.

Tsukuda *et al.*¹ show that double-strand breaks do indeed cause a loss of nucleosomes from sequences within a few thousand base pairs of the break, with kinetics that coincide with those of the loading of Rad51 on to single-stranded DNA. The loss of nucleosomes was catalysed by INO80 and facilitated

by a DNA-repair complex known as Mre11–Rad50–Xrs2, which is rapidly recruited to the break. Furthermore, in yeast strains with mutant Mre11 or INO80 the nucleosomes were not evicted, impairing the subsequent loading of Rad51 on to the DNA. Similarly, in mammalian cells, access of Rad51 to DNA breaks is facilitated by localized chromatin unwinding, which is mediated by the addition of acetyl groups to histones⁶.

In addition to chromatin remodelling next to break sites, a histone protein known as H2AX is phosphorylated very rapidly in a large region surrounding the lesion⁷. Phosphorylation occurs at a conserved motif in the carboxy-terminal tail of H2AX. In budding yeast, phosphorylated H2AX (γ H2AX) is found at least 25 kilobases to each side of a double-strand break induced in the genome, but is depleted within one to two kilobases of the break⁸.

Although phosphorylation of H2AX precedes nucleosome eviction, Tsukuda *et al.*¹ found that it is dispensable for the local displacement of nucleosomes. So, what is the significance of the widespread chromatin modification distal to the break? Previous studies show that γ H2AX is required for the large-scale accumulation of checkpoint proteins to the region around a break, and that loss of H2AX results in defective DNA repair and checkpoint signalling⁹.

Keogh *et al.*² have discovered another function of γ H2AX by asking how it reverts to its unphosphorylated state in budding yeast. Near the break, γ H2AX is rapidly depleted, but farther away (5–20 kb) the γ H2AX is lost only after the binding of the broken strands with the intact donor template. The authors found that a complex (HTP-C) containing the Pph3 phosphatase enzyme specifically regulates the dephosphorylation of γ H2AX; and it seems that the mammalian relative of Pph3, protein phosphatase 2A (PP2A), down-regulates γ H2AX similarly¹⁰.

Despite having high constitutive levels of γ H2AX during normal replication, yeast cells deficient in Pph3 were not growth retarded². Moreover, they could trigger the DNA-damage checkpoint and repair DNA breaks. Nevertheless, the activation of Rad53 and Rad9 — the signalling molecules that initiate the checkpoint — persisted much longer than in normal yeast strains, even after the bulk of DNA damage was repaired. Consequently, checkpoint recovery in Pph3-deficient cells was profoundly retarded. Notably, failure to inactivate the checkpoint was a direct consequence of persistent H2AX phosphorylation because elimination of both γ H2AX and Pph3 led to normal recovery kinetics and inactivation of Rad53 and Rad9.

How does the formation of γ H2AX over a large chromatin domain keep the DNA-damage checkpoint active? One possibility is that DNA-damage response factors may remain activated and bound to the chromatin as long

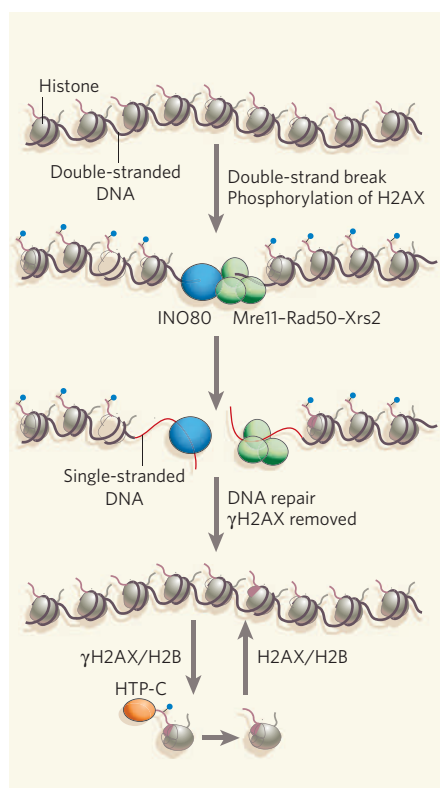


Figure 1 | Chromatin remodelling around double-stranded DNA breaks. Tsukuda *et al.*¹ find that the Mre11–Rad50–Xrs2 DNA-repair complex and the INO80 chromatin remodelling complex are rapidly recruited to breaks and evict nucleosomes in the immediate vicinity of the break. Meanwhile, histone H2AX in a large region around the break is phosphorylated (forming γ H2AX, blue dots). Exonucleases chew back one strand of the DNA, forming a single-stranded overhang (red). Before completion of repair by recombination, an unknown remodelling complex mediates the removal of γ H2AX from 50 kilobases of chromatin; Keogh *et al.*² show that subsequently γ H2AX is dephosphorylated by a protein complex containing the enzyme Pph3, an event that is crucial for release from the damage checkpoint. It is unclear whether entire nucleosomes (shown) or just certain histones (γ H2AX/H2B dimers) are displaced.

as γ H2AX persists. Consistent with this model, in mice that lack H2AX, checkpoint proteins fail to sustain their interaction with DNA breaks at later time points after DNA damage¹¹, and cells from these mice exhibit a defect in the checkpoint that acts between the G2 and M phases of the cell cycle⁹. Keogh *et al.* similarly found that H2AX-deficient yeast deactivated Rad53 and escaped the arrest more rapidly than normal cells². Reversal of H2AX phosphorylation during DNA repair might disrupt associations between checkpoint proteins and the chromatin surrounding DNA damage, interrupting the checkpoint-signalling cascade.

A twist in the tale is that in yeast lacking Pph3, H2AX remains phosphorylated and yet is still displaced from the entire 50-kilobase

region around the break as repair occurs. Remarkably, the checkpoint arrest signal is maintained in these yeast even once the break is repaired and the γ H2AX molecules are redistributed within the nucleus. If the displaced γ H2AX primarily remains soluble, this would mean that DNA-damage signalling can occur 'off' the chromatin. So, at least in budding yeast, dephosphorylation of γ H2AX might be happening after its removal from the chromatin (Fig. 1).

In mammals, it is possible that the attenuation of H2AX may be mechanistically distinct, because the dephosphorylating enzyme PP2A seems to bind to γ H2AX on damaged chromatin¹⁰, and γ H2AX is normally protected from dephosphorylation through interactions with its binding partner MDC1 (ref. 12), a protein that has no direct counterpart in yeast. There may also be other mechanisms of removing γ H2AX, for instance the Tip60 complex in the fruitfly *Drosophila* acetylates phospho-H2Av (the *Drosophila* relative of H2AX), leading to its exchange with unphosphorylated H2Av¹³.

The phosphorylation of H2AX is among the earliest responses to DNA damage, and controls the widespread accumulation of checkpoint response proteins to large chromatin regions surrounding break sites. Nevertheless, it is not required for the rapid recruitment of DNA-repair factors and chromatin remodelling complexes to the actual site of DNA damage¹¹. Tsukuda *et al.*¹ and Keogh *et al.*² demonstrate that dephosphorylation of γ H2AX and its ejection from chromatin regions distal to the break site are crucial in restarting the cell cycle. In this way, the addition and removal of the phosphate group on the tail of H2AX creates a molecular switch that helps to maintain genomic stability.

André Nussenzweig is in the Experimental Immunology Branch, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mail: andre_nussenzweig@nih.gov

Tanya Paull is in the Department of Molecular Genetics and Microbiology, University of Texas at Austin, 1 University Station A5000, Austin, Texas 78712, USA.
e-mail: tpau@icmb.utexas.edu

1. Tsukuda, T., Fleming, A. B., Nickoloff, J. A. & Osley, M. A. *Nature* **438**, 379–383 (2005).
2. Keogh, M. C. *et al.* *Nature* **439**, 497–501 (2006).
3. van Attikum, H., Fritsch, O., Hohn, B. & Gasser, S. M. *Cell* **119**, 777–788 (2004).
4. Morrison, A. J. *et al.* *Cell* **119**, 767–775 (2004).
5. Downs, J. A. *et al.* *Mol. Cell* **16**, 979–990 (2004).
6. Murr, R. *et al.* *Nature Cell Biol.* **8**, 91–99 (2006).
7. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. *J. Biol. Chem.* **273**, 5858–5868 (1998).
8. Shroff, R. *et al.* *Curr. Biol.* **14**, 1703–1711 (2004).
9. Fernandez-Capetillo, O., Lee, A., Nussenzweig, M. & Nussenzweig, A. *DNA Repair (Amst.)* **3**, 959–967 (2004).
10. Chowdhury, D. *et al.* *Mol. Cell* **20**, 801–809 (2005).
11. Celeste, A. *et al.* *Nature Cell Biol.* **5**, 675–679 (2003).
12. Stucki, M. *et al.* *Cell* **123**, 1213–1216 (2005).
13. Kusch, T. *et al.* *Science* **306**, 2084–2087 (2004).

BRIEF COMMUNICATIONS

Molecular trails from hitch-hiking snails

Migrating birds may have transported the *Balea* land snail across vast distances to remote islands.

Darwin was fascinated by the transportation of land snails across great swathes of open ocean by birds — he even immersed snails in sea water to see how long they would survive¹. Here we follow a molecular phylogenetic trail that reveals the incredible transequatorial dispersal of the land snail *Balea* from Europe to the Azores and the Tristan da Cunha islands, and back again. This long-distance dispersal is unexpected for what are proverbially considered the most pedestrian of creatures.

An allegorical picture² from the sixteenth century shows a bird carrying a snail (Fig. 1), but it was not until 1921 that the possibility of land-snail dispersal by birds was discussed for *Balea*³. It was suggested that the exceptionally tenacious slime and arboreal habit of *Balea perversa*, an ovoviviparous hermaphrodite, could aid passive dispersal by birds³. As land snails occur on even the most remote islands, this must be a frequent event, despite the limited number of well documented cases⁴.

In 1824, John Gray assigned two new species of land snail from the Tristan da Cunha islands

to *Balea*, a genus of Clausiliidae thought to be restricted to the Palearctic region, which includes northern Africa and Eurasia to the north of the Himalayas⁵. The Tristan archipelago consists of three main islands situated midway between South Africa and South America at about 37° S, with Gough Island lying about 350 km to the south-southeast (Fig. 2a). It lies 9,000 km and 8,500 km from the Azores and continental Europe, respectively.

On the basis of their extreme biogeographical disjunction, these Tristan land-snail species were subsequently transferred to a new genus, *Tristania*⁶. Additional species of *Tristania* were later described from other islands in the Tristan/Gough group^{7,8}. To reach Tristan from either Europe or the Azores, where two *Balea* species occur⁹, a hostile gap of nearly 9,000 km has to be bridged, which hardly seems possible for a vulnerable pulmonate.

Despite renewed claims for the generic distinctness of *Tristania*, including the erection of the subfamily *Tristaniinae*¹⁰, anatomical evidence indicates that *Tristania* and *Balea* belong to the same genus⁸. We have analysed mitochondrial DNA sequence data for subunit I of cytochrome oxidase for different *Balea* species and find that they confirm this view: our neighbour-joining phylogenetic tree (Fig. 2b, and see supplementary information) indicates that Azorean and Tristan groups of *Balea* arose from a single ancestral species.

As the Baleinae originally radiated in the western Palearctic region^{11,12}, the most parsimonious scenario is that first the ancestral species reached the Azores, then, after additional long-distance dispersal, radiations resulted in two species in the Azores and at least eight in the Tristan/Gough archipelago. One of the Azorean species, *B. heydeni*, reached Madeira and then 'returned' to western



Figure 1 | Etching from Marcus Gheeraerts' fable 'Pride comes before a fall'. An eagle is depicted here, but is unlikely to be the bird that transported *Balea* (see supplementary information). Waders or other migratory birds, which are regular vagrants on mid-Atlantic islands, are more likely vectors.

Europe, where it exists as a distinct species only recently distinguished from *B. perversa*¹³ (the European *B. perversa* reached Iceland relatively late and remains indistinguishable from conspecific European populations).

Edmund Gittenberger*, Dick S. J. Groenenberg*, Bas Kokshoorn†, Richard C. Preece‡

*National Museum of Natural History Naturalis, PO Box 9517, and †Institute of Biology, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands

‡Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
e-mail: rcp1001@cam.ac.uk

1. Darwin, C. *On the Origin of Species* (Murray, London, 1859).
2. Gheeraerts, M. *De Warachtighe Fabulen der Dieren* (Pieter de Clerck, Brugge, 1567).
3. Cockerell, T. D. A. *Nature* **108**, 496–497 (1921).
4. Rees, W. J. *Proc. Malac. Soc. Lond.* **36**, 269–282 (1965).
5. Gray, J. E. *Zool. J.* **1**, 61–62 (1824).
6. Boettger, O. *Ber. Thät. Offenb. Ver. Naturk.* **17/18**, 18–104 (1878).
7. Odhner, N. H. *Proc. Malac. Soc. Lond.* **34**, 168–173 (1960).
8. Preece, R. C. & Gittenberger, E. *J. Moll. Stud.* **69**, 329–348 (2003).
9. Backhuys, W. *Land and Freshwater Molluscs of the Azores* (Backhuys & Meesters, Amsterdam, 1975).
10. Schileyko, A. A. *Ruthenica* **2** (suppl.), 437–564 (1999).
11. Nordsieck, H. *Arch. Molluskenk.* **109**, 67–89 (1978).
12. Nordsieck, H. *Arch. Molluskenk.* **109**, 249–275 (1979).
13. Gittenberger, E., Preece, R. C. & Ripken, T. E. J. *J. Conch.* (submitted).

Supplementary information accompanies this communication on Nature's website.

Received 2 August; accepted 22 December 2005.

Competing financial interests: declared none.

doi:10.1038/439409a

BRIEF COMMUNICATIONS ARISING online
www.nature.com/bca see Nature contents.

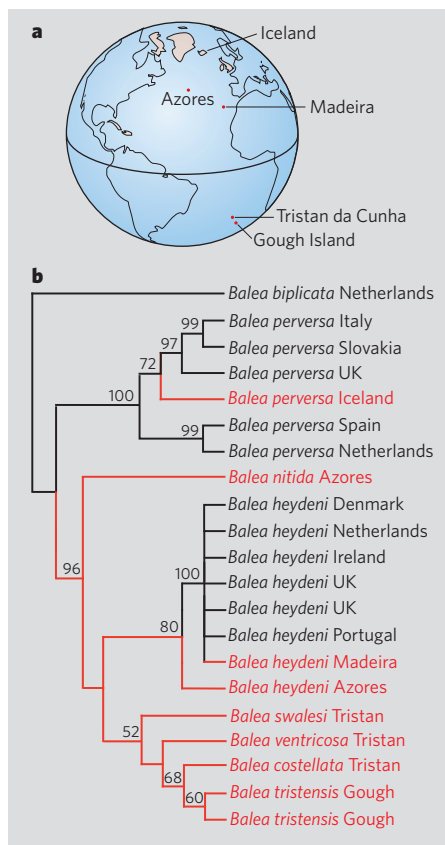


Figure 2 | Distribution of *Balea* species. **a**, Areas where the snails are found. **b**, Neighbour-joining tree for DNA sequences of cytochrome oxidase subunit I in selected species, with bootstrap support values (1,000 replicates). A smaller International Transcribed Spacer 1 data set corroborates these results (see supplementary information). Lineages from mid-Atlantic islands are in red. The extent of sequence divergence between *B. perversa* and its sister clade, together with the occurrence of at least eight *Balea* species on Tristan/Gough⁸, rule out human introduction of the snails. The sequences of populations of *B. heydeni* from mainland Europe are extremely similar, suggesting a recent return to the continent.

EARTH SCIENCE

A wet mantle conductor?

Arising from: X. Huang, Y. Xu & S. Karato *Nature* **434**, 746–749 (2005)

The suggestion that the transition zone of Earth's mantle (410–670 km in depth) is enriched in water is of great possible significance to the geodynamics and geochemistry of Earth's interior, as well as for the role of the mantle in the global water cycle¹. Huang *et al.*² compare the effect of water on electrical conductivities of transition-zone phases to electromagnetism and magnetotelluric soundings of the mantle beneath the North Pacific³ and conclude that the transition zone contains between 1,000 and 2,000 p.p.m. of water, which is considerably more than the 50–200 p.p.m. present in the upper mantle^{4–6}. This conclusion is predicated on the assumption that the transition zone is relatively oxidized, but in fact fairly reduced conditions are more likely^{7,8}. Here I show that if the transition zone is reduced, high conductivities can be explained without the requirement for large enrichments of water.

The experiments of Huang *et al.*² were done in the presence of metallic molybdenum (Mo), meaning that the oxygen fugacity was at or below the buffer defined by coexistence of Mo and MoO₂, which produces an oxygen fugacity similar to that defined by the coexistence of iron metal and wüstite (Fe_{1-x}O). However, Huang *et al.*² recognize that the electrical conductivity of mantle minerals may be critically dependent on oxygen state and assume that the oxygen fugacity of the transition zone of the Earth is similar to conditions imposed by the coexistence of nickel metal and NiO, which

is about four orders of magnitude more oxidizing than their experiments. Accordingly, they correct their measurements to account for this difference, using a plausible but unverified model for the relation between conductivity and oxygen fugacity in hydrous wadsleyite and ringwoodite. In contrast, the high propensity of transition-zone silicates, wadsleyite, ringwoodite and majorite to incorporate Fe³⁺ suggests that Earth's transition zone is quite reduced, with oxygen fugacities similar to or lower than the iron–wüstite buffer^{7,8}. Therefore, it is plausible that the conductivity measurements made by Huang *et al.*² may be applied to the transition zone without significant correction for oxygen fugacity.

Figure 1 compares inverted conductivity profiles for the North Pacific³ with model profiles constructed from equation (1) of Huang *et al.*², which is derived from uncorrected measurements. Reasonable matches to the inverted profiles are found for 100–300 p.p.m. of water. Such values for the water content are comparable to those in the upper mantle that have been deduced from the geochemistry of mid-ocean-ridge basalts^{4–6}.

The conductivity models in Fig. 1 and in Huang *et al.*² are not corrected for the effects of

grain size or secondary phases. Consequently, some uncertainty remains in relating laboratory measurements to the inverted conductivity structure of the mantle transition zone. But if the measurements of Huang *et al.*² give a good approximation of the effect of water on the conductivity in Earth's transition zone, and unless oxygen fugacity at these depths^{7,8} has been grossly underestimated, the available data do not appear to require significant enrichment of water in the transition zone.

Marc Hirschmann

Department of Geology and Geophysics,
University of Minnesota, Minneapolis,
Minnesota 55455, USA

e-mail: marc.m.hirschmann-1@umn.edu

1. Bercovici, D. & Karato, S. *Nature* **425**, 39–44 (2003).
2. Huang, X., Xu, Y. & Karato, S. *Nature* **434**, 746–749 (2005).
3. Utada, H., Koyama, T., Shimizu, H. & Chave, A. D. *Geophys. Res. Lett.* **30**, 1194–1198 (2003).
4. Jambon, A. *Rev. Mineral.* **30**, 479–517 (1994).
5. Michael, P. J. *Earth Planet. Sci. Lett.* **131**, 301–320 (1995).
6. Saal, A. E., Hauri, E. H., Langmuir, C. H. & Perfit, M. R. *Nature* **419**, 451–455 (2002).
7. McCammon, C. *Science* **308**, 807–808 (2005).
8. O'Neill, H. S. C. *et al.* in *Evolution of Earth and Planets* (eds Takahashi, E., Jeanloz, R. & Rubie, D.) 73–88 (AGU, Washington DC, 1993).

doi:10.1038/nature04528

EARTH SCIENCE

Huang et al. reply

Replying to: M. Hirschmann *Nature* **439**, doi:10.1038/nature04528 (2005)

Inference of the spatial distribution of water content in the mantle is critical to our understanding of the dynamics of Earth's interior. A model¹ has been described that indicates there may be a jump in water content at the 410-km discontinuity in the Earth's mantle. From the electrical conductivity, we have inferred² the water content in the transition zone and concluded that it is significantly larger than that of the upper mantle. Hirschmann³ questions our conclusion on the grounds of the trade-off between water content and oxygen fugacity.

We² and Hirschmann³ compare the water content of the upper mantle with that of the transition zone but estimate these by different methods, which introduces errors that are difficult to evaluate. We can also use an alternative approach, in which the water content of the upper mantle and of the transition zone is estimated using the same method. Experimental data on electrical conductivity in

olivine⁴ and wadsleyite are compared with geophysical measurements of the electrical-conductivity jump at 410 km depth (ref. 5). Using the model of Huang *et al.*², a jump in electrical conductivity at 410 km can be expressed in terms of a jump in oxygen fugacity and water content as

$$\frac{\sigma^{410+}}{\sigma^{410-}} = \left(\frac{\sigma_{\text{wad}}}{\sigma_{\text{oli}}} \right)_0 \left(\frac{f_{\text{O}_2}^{410+}}{f_{\text{O}_2}^{410-}} \right)^{-1/8} \left(\frac{C_{\text{W}}^{410+}}{C_{\text{W}}^{410-}} \right)^{3/4}$$

where $\sigma^{410+}/\sigma^{410-}$ is the conductivity ratio (below and above 410 km, which is about 10; ref. 5), $(\sigma_{\text{wad}}/\sigma_{\text{oli}})_0$ is the conductivity ratio of wadsleyite and olivine compared at the same condition, $f_{\text{O}_2}^{410+}/f_{\text{O}_2}^{410-}$ is the oxygen fugacity ratio, and $C_{\text{W}}^{410+}/C_{\text{W}}^{410-}$ is the water-content ratio. On the basis of our results² and a re-analysis of those on olivine⁴, we can now place tight constraints on $(\sigma_{\text{wad}}/\sigma_{\text{oli}})_0$.

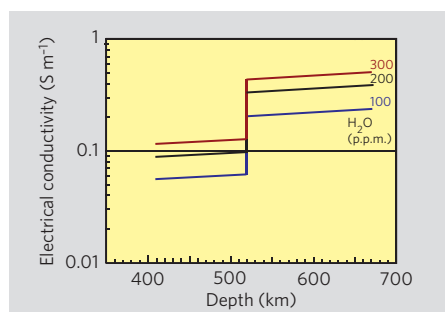


Figure 1 | Electrical conductivity versus depth in Earth's transition zone. Lines are for wadsleyite (410–520 km) and ringwoodite (520–670 km) calculated from equation (1) and the parameters in Table 1 of Huang *et al.*² for 100, 200 and 300 parts per million (p.p.m.) H₂O. The temperature is assumed to increase from 1,825 to 1,900 K between 410 and 670 km. The grey shaded region represents the range of conductivities deduced from three different inversions of conductivity profiles of geoelectrical sounding and magnetotelluric response function data for the North Pacific³. The conductivity profiles derived from experimental data are appropriate for oxygen fugacities similar to the Fe–Fe_{1-x}O buffer.

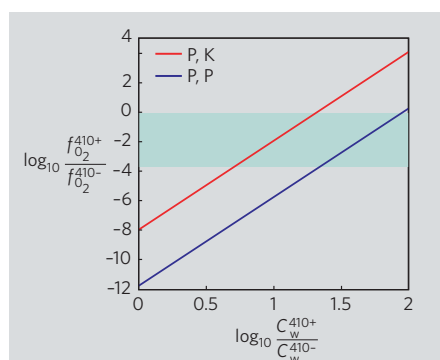


Figure 1 | Trade-off between water-content ratio and oxygen-fugacity ratio. The trade-off is consistent with the geomagnetic observation of a jump in conductivity at the 410-km discontinuity by a factor of about ten. Red line corresponds to a case where the Paterson calibration⁶ (P) for water content is used for wadsleyite, but the Koga *et al.* calibration⁷ (K) is used for olivine; the blue line corresponds to a case where the Paterson calibration⁶ is used for both minerals. Turquoise shaded region, plausible range of oxygen-fugacity ratio between the transition zone and the upper mantle that corresponds to the oxygen fugacity for the Ni–NiO and Mo–MoO₂ buffers.

These two different data sets were obtained using the same technique (sample assembly, including the control of oxygen fugacity) and nearly identical grain size, so the results can be compared with little ambiguity. Note that the earlier report of high conductivity in wadsleyite² is almost entirely due to the high water content. If the conductivity of olivine and wadsleyite is compared under the same conditions (water content, temperature, pressure, and so on), the difference is very small: $(\sigma_{\text{wad}}/\sigma_{\text{oli}})_0 \approx 0.3\text{--}1$ at the 410-km condition, depending on the water-content calibration.

Using these values, the range of $f_{\text{O}_2}^{410+}/f_{\text{O}_2}^{410-}$ and $C_{\text{W}}^{410+}/C_{\text{W}}^{410-}$ can be calculated, which is consistent with geophysical observation (Fig. 1). It is evident that if all the observed jump in conductivity were due to the difference in oxygen fugacity, without any jump in water content, then a reduction in oxygen fugacity of 8–12 orders of magnitude would be needed across the 410-km discontinuity, which is unacceptable. Within a plausible range of oxygen fugacity, we infer that the increase in water content is about one order of magnitude across the discontinuity to explain

the observed jump in electrical conductivity.

In summary, a new analysis of the electrical-conductivity jump at 410-km depth provides a robust estimate of a jump in water content at 410 km by a factor of about ten, in agreement with ref. 1.

Xiaoge Huang*†, **Yousheng Xu***,
Shun-ichiro Karato*

*Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520, USA
e-mail: shun-ichiro.karato@yale.edu

†Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China

1. Bercovici, D. & Karato, S. *Nature* **425**, 39–44 (2003).
2. Huang, X., Xu, Y. & Karato, S. *Nature* **434**, 746–749 (2005).
3. Hirschmann, M. *Nature* **439**, 10.1038/nature04528 (2005).
4. Xu, Y., Poe, B. T., Shankland, T. J. & Rubie, D. C. *Science* **280**, 1415–1418 (1998).
5. Utada, H., Koyama, T., Shimizu, H. & Chave, A. D. *Geophys. Res. Lett.* **30**, 10.1029/2002GL016092 (2003).
6. Paterson, M. S. *Bull. Mineral.* **105**, 20–29 (1982).
7. Koga, K., Hauri, E. H., Hirschmann, M. M. & Bell, D. R. *Geochem. Geophys. Geosyst.* **4**, 10.1029/2002GC000378 (2003).

doi:10.1038/nature04529

REVIEWS

The search for a topographic signature of life

William E. Dietrich¹ & J. Taylor Perron¹

Landscapes are shaped by the uplift, deformation and breakdown of bedrock and the erosion, transport and deposition of sediment. Life is important in all of these processes. Over short timescales, the impact of life is quite apparent: rock weathering, soil formation and erosion, slope stability and river dynamics are directly influenced by biotic processes that mediate chemical reactions, dilate soil, disrupt the ground surface and add strength with a weave of roots. Over geologic time, biotic effects are less obvious but equally important: biota affect climate, and climatic conditions dictate the mechanisms and rates of erosion that control topographic evolution. Apart from the obvious influence of humans, does the resulting landscape bear an unmistakable stamp of life? The influence of life on topography is a topic that has remained largely unexplored. Erosion laws that explicitly include biotic effects are needed to explore how intrinsically small-scale biotic processes can influence the form of entire landscapes, and to determine whether these processes create a distinctive topography.

Do biota affect landscape form and evolution? One way to think about this question is to imagine a very high-resolution topographic map of the Earth (for example, 1 m^{-2} data density) from which all artefacts of human activity have been removed and all vegetation has been cleared, and consider if a unique topographic signature of life would be evident in such a map. In other words, has the emergence of life so fundamentally altered erosion mechanics that there are features of the landscape that could only exist owing to biotic influences—if life had not arisen, would the tectonic and climatic processes that drive uplift and erosion of landscapes be significantly different? The simple question about the effect of biota on landscape development therefore raises several essential issues about the mechanisms underlying Earth's evolution.

Anyone who has tended a garden knows that vegetation affects erosion processes. Remove it, baring the soil, and surface erosion is likely. Much practical work has been done to quantify the influence of biota on erosion¹. But over the long geologic timescales of landscape evolution, as mountain ranges emerge and climb towards an erosional equilibrium with the mass flux due to colliding plates, are biogenic processes significant? Would the Appalachians, Alps, Andes or Himalayas have distinctly different topography in the absence of biotic processes? Would the deeply weathered landscapes of Australia, South America or Africa have formed in the absence of life? Are river networks, now widely recognized as large-scale examples of self-organization², influenced by biotic processes? Despite the fact that all current landscapes emerged in a biotic world, few landscape evolution models explicitly include the effects of life^{3–7}.

In this Review, we explore these questions in three ways. First, we consider how biotic processes influence weathering, erosion, and sediment transport mechanisms, and how such influences could be manifest in landscape-scale morphology. Second, we consider how biotic processes affect climate and tectonics, which drive landscape evolution. Last, we consider landscapes on an Earth on which life never arose and examine the topographies of Venus and Mars.

Erosion, biotic processes and landscapes

Recently, numerical modelling has become an important tool in the study of landscape form and evolution. Numerical models have been used to investigate the process controls on landscape form^{8–10}, the coupling of tectonics and erosion^{11–14}, and the relationships among climate, tectonics and topography^{14,15}. Some studies have specifically explored the effects of vegetation on landscape processes and form^{3–7}. These modelling studies have provided valuable new insight into landscape dynamics, but their applicability is limited by a lack of mechanistic mathematical expressions for modelling specific erosion processes. All landscapes must obey an equation for the conservation of mass^{16,17}:

$$\frac{\partial z}{\partial t} = U - E - \nabla \cdot \mathbf{q}_s \quad (1)$$

in which z is the elevation of the ground surface, t is time, U is uplift rate, E is the incision rate into bedrock, and $\mathbf{q}_s$ is the volume flux of stored sediment (soil, colluvium, alluvium, and so on) per unit width. This is a simplified form of the conservation equation in which changes in bulk density have been ignored and, correspondingly, solute losses are not considered. To solve equation (1), we require an understanding of the tectonic processes (U) operating on the landscape as well as 'geomorphic transport laws'¹⁷ (E and $\mathbf{q}_s$), which describe the rates of different transport, bedrock-to-soil conversion, and erosion processes in terms of material properties, climatic influences and attributes of the topography and subsurface. In this Review, we make the case that all three terms on the right-hand side of equation (1) are influenced by biota.

To understand the role of transport laws in shaping landscapes, look out of any aeroplane window as you travel across the world. What you will most commonly see is a landscape composed of ridges and valleys, forming a hierarchical structure in which smaller valleys drain to successively larger ones (Fig. 1). Why are there ridges and valleys? What controls their size? What sets the height and steepness of the hills?

The seminal paper by Smith and Bretherton¹⁸ took an important step towards answering these questions by showing how two simple

¹Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA.

approximations for geomorphic transport laws (q_s in equation (1)—in this simple case, the bedrock erosion term E is ignored) give rise to either ridges or valleys (Fig. 1). For each transport law, they first solved equation (1) for the steady-state case ($\partial z/\partial t = 0$) and then introduced lateral perturbations to the topography. If $q_s = KS$, where K is a diffusion-like coefficient and S is the local topographic slope, the resulting steady-state hillslope is convex-up (Fig. 1). Lateral topographic perturbations to this surface are damped by the diffusive sediment transport. Now consider the case in which sediment transport is proportional to both local water discharge (q_w) and slope, that is, $q_s = F(q_w, S)$. Because the water discharge at a point on the landscape increases with the size of the area that drains to that point (A), gentler slopes are sufficient to transport sediment further from the topographic divide, and the resulting steady-state landform is concave-up (Fig. 1). Perturbations to this surface are unstable, however, because local concavities cause an increase in drainage area, which in turn accelerates incision and valley growth. Thus, ridge-and-valley topography emerges from a competition between diffusion-like and advective erosion processes. If we replace $F(q_w, S)$ with $cA^m S^n$ (which assumes that incision varies with runoff energy expenditure or flow shear stress), with c , m and n empirical parameters, we can write in general $q_s = KS + cA^m S^n$ and solve

equation (1) to produce ridge and valley topography. This is, in essence, what nearly all landscape evolution models currently do. But the two proposed transport laws are not deeply mechanistic. What are the mechanics of erosion, and how does life fit in?

Recent reviews^{17,19,20} discuss the current state of knowledge about geomorphic transport laws, so we can summarize here and focus on the biotic influence. Table 1 lists the primary erosion processes driving landscape evolution, the corresponding geomorphic transport laws, and the abiotic and biotic mechanisms that influence those laws. Here we discuss some of the critical processes in more detail.

Massive bedrock must break down into smaller fragments before it can be eroded and transported away. Abiotic processes such as chemical weathering, salt growth, freezing, stress release fracturing and chemical dissolution can produce loosened particles. So, too, can bedrock landslides. Field studies^{17,20,21} have shown, however, that in a wide range of soil-mantled landscapes, soil production (the production of loose, transportable material from bedrock) is primarily due to biogenic disturbance. Root growth and animal burrowing disrupt bedrock that is weathered but structurally intact, creating a loose material free to move downslope²⁰. The probability of biotic contact with the rock should decline with increasing soil mantle thickness, and this expectation is borne out in field studies of soil production rates. Using cosmogenic radionuclide dating, Heimsath *et al.*²¹ documented an exponential decline in the rate of soil production with depth ($P = P_0 e^{-\alpha h}$, in which P is the production rate with units of $L T^{-1}$, h is soil depth, P_0 is the production rate on exposed bedrock, and α is a parameter with units of L^{-1} ; note that P is part of E in equation (1)). The production rate of soil is also strongly influenced by the degree of weathering of the bedrock; and biotic processes, especially microbial activity, can strongly affect weathering rates^{22,23}. Where soil erosion outstrips soil production, bedrock emerges and the role of biota is greatly diminished.

Granular soils on inclined surfaces experience net downslope movement when they are dilationally disturbed²⁴. The source of disturbance can be abiotic, as in the case of freeze-thaw, wetting-drying, or salt growth processes. Some clay-rich soils can swell and flow downslope²⁵. On soil-mantled landscapes, however, dilational disturbance by burrowing organisms and displacement by falling trees often appears to dominate motion^{24,26}. This can lead to a linear or nonlinear dependence of soil flux on local slope^{27,28}. Despite the observation that biotic processes may dominate the downslope movement of soil, there are currently no data available to define quantitative relationships between flux law parameters and specific biotic processes (although significant work has begun²⁰). Most commonly it is assumed that the proportionality constant between flux and slope (K or K_{nl} in Table 1) depends in some way on biotic activity.

In semi-arid to arid landscapes, where vegetation is sparse, precipitation rates can exceed the meagre infiltration rate into unvegetated ground and generate overland flow and surface wash. Where vegetation is sufficiently dense, it dictates the rate and pattern of surface wash and strongly limits its efficacy¹. Currently there are no geomorphic transport laws that describe these processes, although empirical expressions for land management do explicitly account for biotic influences on surface wash¹. Nonetheless, we can reason that in a wet abiotic world, loose particles on hillslopes would be washed away, hindering the development of a soil mantle.

River incision into bedrock drives landscape evolution, and thus a geomorphic transport law for this process is crucial to exploring linkages between tectonics, climate and erosion. Over the past two decades, theory, experiments and field studies have shown that an initial hypothesis^{29–31}—that the rate of erosion is proportional to the rate of energy expenditure—while very useful¹⁹, is incomplete. Recent work emphasizes the role of sediment grain size and quantity in controlling incision rates^{19,32,33}. The sediment load carried along the bed of a river acts as an abrasive tool (more of it leads to more wear) as well as a cover (the more of it, the more protected the bed). The

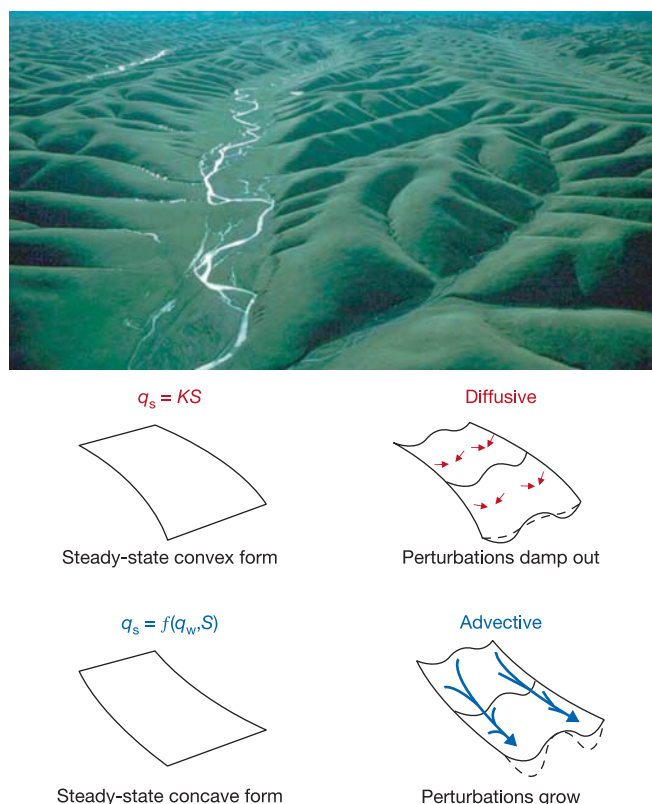


Figure 1 | An explanation for the origin of ridges and valleys. Top, photograph of a grass-covered, soil-mantled landscape finely divided into ridge-and-valley topography near Orland, California. Photo courtesy of J. Kirchner. Bottom, an explanation, first proposed by Smith and Bretherton¹⁸, of why such topography emerges during landscape evolution (see text for a summary of their analysis). Slope-dependent (diffusive) transport leads to convex hillslopes, and when the topography is laterally perturbed, the transport direction (red arrows) causes the high points to lower and low points to fill in. This results in smooth topography, as suggested by the dashed line. In contrast, transport that depends on water flow and slope advects sediment, and produces concave-up hillslopes. Flow concentrations (blue flow paths) resulting from lateral topographic perturbation lead to incision, as suggested by the dashed lines. The competition of these two processes leads to diffusion-dominated ridges and advection-dominated valleys.

larger the rocks entering the channel from adjacent hillslopes, the steeper the channel must become to transmit them downstream. Biotic processes generally reduce the size of the grains entering the channel. Biotic effects tend to increase the weathered state of the rock, which accelerates the breakdown of grains once they enter the channel. Weathering of the bedrock beds of rivers (perhaps facilitated by biotic processes) would also increase their erodibility. Other primary erosion processes (landsliding, debris flows, glaciers and wind) and the associated biotic effects are summarized in Table 1.

As indicated in Table 1, then, biotic processes strongly influence the production and transport of debris, yet the few geomorphic transport laws that have been proposed do not account explicitly for these processes. Landscape evolution models that explore the relationships among climate, topography and tectonics rely on these transport laws. The absence of transport laws for many processes and the lack of accounting for biotic effects make the interpretations derived from these models less reliable. Nonetheless, we have sufficient understanding to explore qualitatively how erosion processes and the resulting landforms might differ in the absence of life.

Landscape evolution in the absence of life

What would happen if all life suddenly disappeared and the artefacts of humankind (roads, dams and so on) had never been constructed

(Fig. 2)? For the moment, we assume that the oceans, Earth's climate and tectonics remain unchanged; we will consider these interactions below. Three distinctive responses would be likely.

The most obvious and certain effect of the sudden absence of life on Earth would be the rapid erosion of the soil mantle that covers semi-arid to wet landscapes. The absence of vegetative root strength would permit landsliding in loose soil on steep slopes during intense rainstorms. Soil consolidation owing to a lack of dilational disturbance by root growth and animal burrowing would lower infiltration rates, promoting overland flow, sheetwash and gully. Rocky scree slopes fringing steep mountains would remain relatively unchanged.

Our first hypothesis, then, is that in an abiotic world, the soil mantle that drapes many hilly and mountainous landscapes would be stripped, and rough, poorly weathered, bedrock surfaces would emerge (Fig. 3a). The deep weathering profiles that develop on bedrock in wet tropical areas would be absent. One might speculate, therefore, that smooth, rounded, soil-mantled hilltops are a topographic signature of a biotic world. The landscape of the Atacama Desert of Chile proves that this is untrue as a universal generalization (Fig. 3b). In the hyperarid inland region of the Atacama referred to as the Central Depression³⁴, there are extensive areas of rounded, soil-mantled topography in which biotic processes are essentially absent. Intensive salt weathering breaks down bedrock to produce soil, and the surface mantle receives an input of wind-blown sediment³⁵. The

Table 1 | A summary of geomorphic transport laws and biotic effects

Process	Geomorphic transport law, GTL*	Abiotic mechanisms	Biotic mechanisms
Soil production rate	$P = P_0 e^{-\alpha h}$ (ref. 21); other expressions proposed in refs 67–69	Salt and freeze–thaw weathering, atmospheric dust input, mineral alteration leading to loss of physical strength	Animal burrowing, root growth and tree throw; microbially mediated geochemical reactions; P_0 (through weathering) and α depend on biota
Slope-dependent downslope movement (creep)	$q_s = -K \nabla z$ (refs 70, 28); $q_s = \frac{-K_n \nabla z}{1 - (\nabla z / S_c)^2}$ (ref. 27); see ref. 20 for equations explicitly including biogenic mechanisms	Wetting and drying, freezing and thawing ^{69,71} , shear flow	Animal burrowing and tree throw that cause dilational disturbance; both K_n and S_c depend on biota
Landsliding	None available, but important starts for earthflows ⁷² , deep-seated landslides ⁷³ and landslide dynamics ⁷⁴	Stress exceeds material strength owing to earthquakes, elevated pore pressures derived from precipitation or from undermining of toe; released sediment travels downslope	Roots add strength and vegetation canopy reduces peak rainfall intensity (especially important in colluvial soil failures); evapotranspiration may reduce water levels in potential deep-seated landslides; vegetation may affect travel distance
Surface wash and splash (Horton mechanisms)	Many short-term empirical and mechanistic expressions ¹ , but no GTL available	Rainsplash and overland flow displace and remove particles; rill and gully incision	Strongly affects all aspects of runoff and erosion processes
River incision into bedrock	$E = k_b A^m S^n$ (refs 17, 19, 29); $E = k_1 \frac{q_{sp}}{(T-1)^{1/2}} - k_2 \frac{q_{sp}^2}{D^{3/2}(T-1)^2}$ (ref. 33)	Plucking and particle wear due to river flow and sediment transport	Indirect effect through influence on sediment grain size and bank strength (which affects channel width); large woody debris may hold sediment and retard incision rate
Debris flow incision into bedrock	$E = k_d f \left[\rho_s D^2 \left(\frac{\partial u}{\partial y} \right)^a L_s \right]^p$ (ref. 75)	Particle impact and sliding wear of bedrock during mass transport	Large woody debris may halt debris flow movement, or enhance erosion
Glacial scour	$E = c U_b$ (ref. 76); more mechanistic expressions proposed ⁷⁷	Sediment-rich basal sliding wears bedrock	No apparent influence
Wind transport and scour	Extensive theory for sediment transport by wind ^{78–80} , some theory for rock abrasion ⁸¹	Abrasion by wind-suspended particles	Biotic ground cover eliminates process ⁸²

Details about each process and transport law can be found in the cited references. See ref. 17 for a discussion of GTLs. Note that none of the proposed GTLs define terms that are specific to biotic mechanisms.

* Approximate definitions of terms (see cited references for details): P , soil production rate from bedrock (contributes to E in equation (1)); P_0 , soil production from exposed rock; h , soil thickness; α , constant; q_s , volumetric sediment transport rate per unit width; K , constant; z , elevation; K_n , constant; S_c , threshold slope at which transport rate becomes infinite; E , incision rate into bedrock; k_b , constant that may depend on uplift rate and rock strength; A , drainage area; S , local slope; m, n , constants; k_1 , constant that depends on bedrock strength; k_2 , constant that depends on bedrock strength and force needed to initiate sediment transport; q_{sp} , bedload supply; T , ratio of shear stress to critical shear stress for initial motion; D , representative grain diameter; k_d , constant that depends on bedrock properties; f , frequency of debris flows; ρ_s , bulk density of debris flow; u , debris flow velocity as a function of distance y above the bed; L_s , length of debris flow 'snout'; a, p , constants; U_b , basal ice velocity; and c , constant.

resulting topography appears almost identical to that developed in the presence of active biotic processes (Fig. 1). This similarity in form is measurable: the curvature (a quantitative measure of hillslope 'roundness', usually expressed as the laplacian of elevation, $\nabla^2 z$) of the hillslope in Fig. 3b is similar to that commonly observed^{21,36,37} in well-vegetated environments (Chilean data; J. Owen, personal communication). We can conclude that, whereas rough bedrock surfaces would be more typical of an abiotic world, rounded, soil-mantled landscapes could also occur. Figure 3a and b also illustrates that exposed bedrock is more likely where hillslopes are bordered by actively incising channels.

The stripping of the soil mantle and the emergence of bedrock on hillslopes should introduce coarser, less weathered sediment to channels. Our second hypothesis, then, is that river beds in an abiotic world would be coarser. This coarsening would cause the channels to steepen, and the longitudinal profiles of rivers would rise to greater heights in the mountains. This implies that the adjacent ridges would also be higher. While the occurrence of steep, rocky channels would increase and mountains might be somewhat higher in an abiotic world, these features are not unique and would not constitute a distinctive biotic signature when compared to the world we live in.

The tendency of river channels to meander (shift laterally along curved paths) or braid (form multiple channel paths) depends in part on the mechanical strength of the river banks, which is influenced by

vegetation. Weak channel banks allow high flows to erode laterally, widening the channel and giving rise to multiple channel paths separated by bars (braiding). This realization has led some to propose that meandering is a signature of the influence of land plants^{38,39}. Our third hypothesis, then, is that meandering rivers would be rare or absent in an abiotic world. While the influence of vegetation on bank strength is well documented^{40,41}, it is important to note that other sources of strength can lead to meandering. Bedrock can provide the appropriate strength⁴², and meandering bedrock canyons are observed on both Earth and Mars⁴³. But what about the case of meandering channels with banks composed of their own alluvial floodplain sediments? Recent observations^{44–46} demonstrate that large-scale meanders with cut-off loops developed on a deltaic fan surface on Mars in the absence of vegetation. The source of strength there may have been frozen ground⁴⁴. It is likely that meandering rivers would be less common on an abiotic Earth, but we cannot conclude that they would be absent.

These three hypotheses have helped us to explore the ways in which landscapes would differ in the absence of biota, but none have led us to identify a landform that cannot exist without life. We now turn our attention to the influence of biota on climate and, consequently, large-scale erosion and tectonics.

Landscapes if life influences climate

Life mediates atmospheric and ocean chemistry. Vegetation, in particular, influences heat transfer, water transport and characteristics of the land surface that affect atmospheric turbulence. How

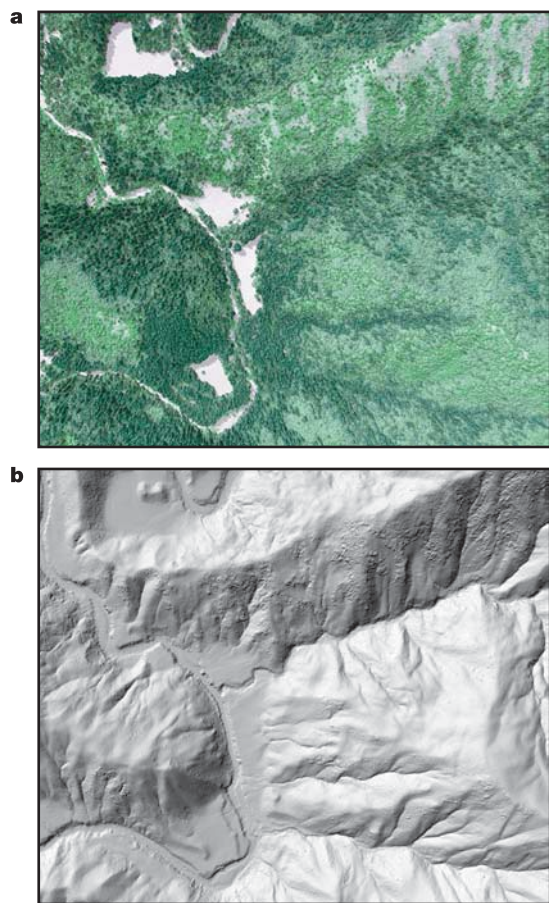


Figure 2 | Shaded relief images derived from airborne laser swath mapping (ALSM) data. Panel **a** includes vegetation, with colours based on canopy heights; panel **b** has been filtered to an approximate bare-earth surface. Width of the images is 2.3 km, and horizontal resolution of the data is about 1 m. ALSM data were collected and processed by the National Center for Airborne Laser Mapping (NCALM). Area shown is in the Angelo Coast Range Reserve along the South Fork Eel River, California (part of the University of California Natural Reserve System and a National Center for Earth-surface Dynamics research area).



Figure 3 | Hillslopes on which soil formation appears to be driven entirely by abiotic processes. **a**, Bedrock hillslope in an area of active channel incision in the hyperarid Atacama Desert, Chile. **b**, Rounded, soil-mantled hillslope in the Atacama Desert bordered by an aggrading plain. Hillslopes in both **a** and **b** are about 40 m high. **c**, Image mosaic (artificially coloured) of the Columbia Hills, Mars, acquired by Spirit, one of the two Mars Exploration Rovers, from positions approximately 300 m from the base of the hills. Once the sky has been artificially coloured blue (the true colour is yellowish grey) and the red colour of the rocks (due to the presence of oxidized iron) has been removed, the rounded, soil-mantled hillslopes in **c** can hardly be distinguished from the terrestrial landscape in **b**. Mars image courtesy of NASA/JPL/Cornell University.

might life shape the pattern of precipitation, which is one of the major drivers of landscape evolution? One way to address this question is to perform a virtual experiment: remove all vegetation from a global climate model and document the effects on the predicted spatial patterns of rainfall and temperature. Kleidon *et al.*⁴⁷ report just such an experiment. A key finding is that the removal of vegetation leads to large changes in mean annual precipitation over extensive geographic regions.

Recent modelling studies, which have given rise to numerous field investigations, suggest that uplift, erosion and climate are strongly coupled in evolving mountain belts^{14,48,49}. Figure 4 illustrates this proposed coupling. Crustal thickening at convergent plate boundaries causes uplift of the land surface, which steepens regional topographic gradients. In response, rivers carve canyons, steepening local slopes and inducing further landscape erosion. This erosional mass removal alters the force balance across the collisional zone and ultimately influences the height to which the mountains climb, the breadth of the uplifted region and the overall shape of the orogen. Through this feedback mechanism, the uneven distribution of precipitation across most mountain ranges exerts an important

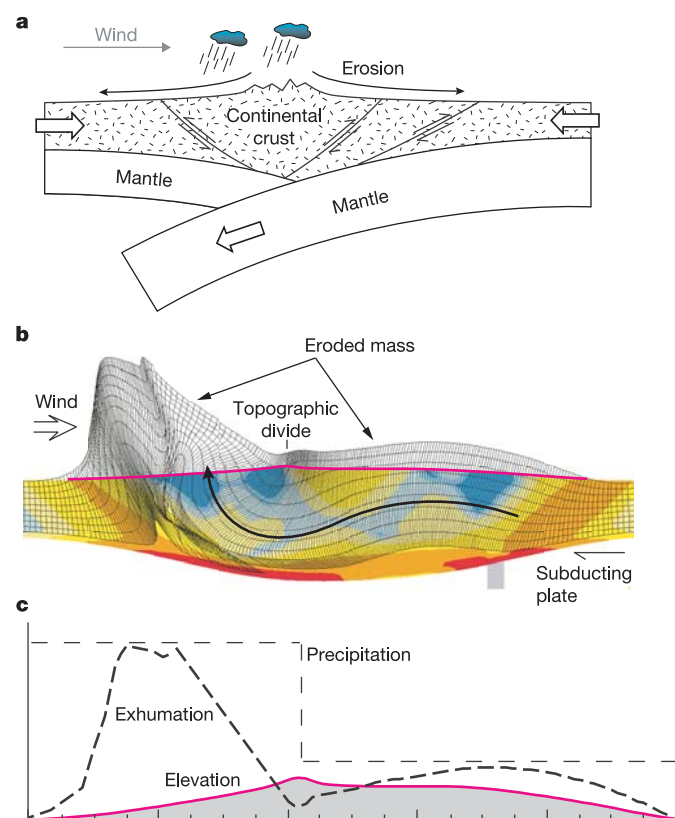


Figure 4 | Links among tectonics, climate, erosion and topography at convergent plate boundaries. **a**, Hypothetical cross-section of convergent plates (composed of continental crust overlying flowing mantle, with the plate on the right subducting beneath that on the left) and the resulting mountain range, as influenced by climatically controlled, spatially varying erosion. Open arrows show direction of motion; smaller coupled arrows show faulting arising from the convergence. **b**, Result of a finite-element numerical model of the scenario in **a** developed under strongly asymmetric rainfall across the mountain range in which wind-driven storms drop more precipitation on the left side of the mountain, leading to greater river incision and landscape erosion. The asymmetric rainfall pattern causes the topographic divide to shift towards the side of the mountain that receives more rainfall. Warmer colours correspond to higher strain rates. The magenta line marks the topographic surface, and the portion of the mesh above this line shows the eroded mass. The curved black arrow shows the path of bedrock through time. **c**, Simplified plot of exhumation, elevation and precipitation for the model result in **b**. Figure modified from Willett¹⁴.

control on the dynamics of mountain growth: if erosion is positively correlated with precipitation, the wetter sides of mountains will erode more rapidly, accelerating the uplift of the underlying bedrock. This analysis suggests that biota, by influencing the pattern of precipitation, affect the height, width and symmetry of mountain ranges. Yet we still cannot argue that there is a unique mark of life upon the landscape.

The numerical experiment of Kleidon *et al.* also predicts changes in near-surface air temperature of ± 4 K, with some high mountain areas (parts of the Andes and Rocky Mountains, for example) experiencing strong cooling. Such a temperature drop would expand the area of freeze–thaw activity, which may increase erosion rates, but the net effect on topography will depend on the rate of debris removal by receiving streams. Again, this effect may lead to a change in the local erosion rate, but it would not make an abiotic Earth topographically distinct.

What if life never arose?

From a landscape evolution perspective, the question to ask is, if life had never emerged on Earth, how would the coupled climate–lithosphere system have evolved? No consensus was found in the literature on this issue; in fact, very little has been written about the evolution of Earth in the absence of life. The role of life in the evolution of the Earth is debated, as summarized in Box 1. To gain some insight into this problem, we can look to other Solar System bodies, particularly our neighbours Venus and Mars.

On Earth, plate tectonics depends on an upper mantle low-viscosity zone on which plates can glide, and it has been proposed that this zone arises from injections of water at subduction zones⁵⁰. This insight suggests that the scarcity of water on the surface and in the interior of Venus may be responsible for the absence of plate tectonics there, despite internal dynamics that triggered a magmatic resurfacing of the planet within the past 700 million years⁵¹. Venus is relatively smooth (80% of the planet lies within 1 km of the mean planetary radius⁵⁰) and lacks any evidence of non-volcanic surface channels. Early Venus would have received roughly the same amount of insolation as the present-day Earth. Is it possible that the emergence of life on Earth prevented the development of atmospheric conditions favourable to the loss of water via solar wind erosion, keeping the planet ‘wet’ and enabling plate tectonics? Is plate tectonics on Earth a consequence of life on Earth? One contributing effect is certainly the biologically mediated sequestration and burial of carbonate in the oceans, which prevents a runaway build-up of atmospheric carbon dioxide like that found on Venus. As noted by Chyba⁵², the current understanding of early Earth’s atmospheric chemistry and the development of life is experiencing a productive turmoil. This will also lead to better insight about how the Earth could have evolved in the absence of life.

Perhaps the greatest surprise to those who have examined the wonderfully detailed images of Mars acquired by robotic orbiters and rovers is how familiar the landscape feels, particularly at the local scale. Channel networks are widespread^{53–55}, river meanders in bedrock canyons and in alluvium have been found, and large alluvial fans have recently been mapped⁵⁶. Rocky ridges similar to the one in Fig. 3a are common on Mars, but convex, debris-mantled hillslopes like the one in Fig. 3b can also be found (Fig. 3c). Wind-formed dunes are a pervasive feature of the landscape, extensive glacial features have been mapped⁵⁷ and numerous volcanoes exist.

Of course, Mars’ topography does look different: impact craters abound, undissected plains are extensive, and there are some features, believed to be glacial in origin, that do not appear to have any analogues on Earth^{58,59}. Mars is a one-plate planet with a very different large-scale structure, such as the enormous Tharsis bulge and the striking hemispheric dichotomy. This suggests that the question on Mars should be turned around: are there any features found there that could not occur in the presence of significant biotic influences? Perhaps some glacial features require ice-forming

conditions incompatible with extensive life, but otherwise we are unable to identify any such features. The surface of Mars contains an abundance of familiar landforms, and we are beginning to see that Mars is not unique among Solar System objects in this respect. It is particularly noteworthy, for example, that the most prominent landscape features revealed during the descent of the Huygens probe to the surface of Titan are valley networks⁶⁰, which dominate the surface of the Earth.

The importance of life in landscape processes

Life permeates the Earth's surface and controls or mediates erosion and transport processes. Biotic processes influence long-term landscape evolution, even to the extent that they affect the height, width and symmetry of mountain ranges. Yet the answer to the question 'Is there a unique topographic signature of life on Earth?' seems to be 'no'. By 'unique' we mean a landform that could only exist in the presence of life. Even a single occurrence of broadly convex hillslopes on Mars (Fig. 3), for example, indicates that this landform is not a unique signature of life. This does not mean that life has not profoundly altered the course of Earth's evolution or its landscapes. Indeed, we expect that there is a topographic signature of life, but it may be more subtle than the presence of certain diagnostic landforms. This subtlety is likely to be one of frequency of occurrence and of scale.

Figure 5 illustrates the issue of frequency of occurrence. If we were to make quantitative measurements of landscape properties (for example, slopes, heights or curvatures) on an abiotic Earth and compare them to measurements of the Earth we know, the analysis presented above suggests that the frequency distributions of these measurements would be very different, even though all observed landform types would be found in both worlds. If we were to walk around an abiotic Earth, it would look different from ours, but it would not exhibit landforms unfamiliar to us, nor would we find any landform missing from our past experience. Mountainous

landscapes would be rocky and irregular, and most rivers would have coarse beds and braided channels, but such features are observed on Earth today. The difference would lie in the frequency distributions of certain landform properties.

What might be the properties to plot on the abscissa in Fig. 5? Process models predict^{9,61}, and empirical analyses of high-resolution topography show¹⁷, that slopes generally steepen with increasing drainage area on hillslopes and become more gentle with increasing drainage area in valleys, with the steepest slopes occurring at the transition from hillslopes to valleys. Analysis of the slope–area relationship has proven useful as a means of examining the results of landscape evolution models^{7–10,61}. The inset in Fig. 5 shows slope–area curves for three landscapes, one in which biota have a significant influence on erosion processes, and two in which the biotic influence on erosion is minimal. These plots suggest four properties of the slope–area relationship whose frequency distributions could be used to distinguish biotic and abiotic landscapes. These properties are the power-law exponents for hillslopes (small areas) and valleys (large areas), and the area and slope at the transition point between hillslopes and valleys. At larger drainage areas and gentler slopes, further changes in erosion processes may produce additional breaks in the slope–area relationship⁶². The slope–area properties of the biotically-influenced landscape in Fig. 5 clearly differ from those of the landscapes with minimal biotic influence. The high-resolution topographic data needed to perform these analyses are currently available for very limited portions of the Earth's surface, but as airborne laser swath mapping surveys expand this coverage, it will

Box 1 | The influence of life on Earth's evolution

When the Solar System was young, conditions at the surface of our planet were very different from what they are today. The Earth was being bombarded by chemically reactive comets and asteroids, and mantle convection was more vigorous while continental crust was accumulating^{83–85}. The chemistries of the mantle, crust and atmosphere were co-evolving^{83–85}. Life appears to have arisen by 3.5 billion years ago (although the earliest dates are controversial^{86,87}) and there is evidence that by this time oceans and sediments had formed^{84,85}, despite the fact that the Sun's luminosity was ~30% lower than its present value. Today, such a large reduction in insolation would cause the oceans to freeze, and so the evidence that the early Earth was not frozen has prompted many proposed solutions to the 'faint young Sun paradox'⁸⁵. It has been argued that before 3.5 Gyr ago, when impacts were common, the Earth was indeed frozen, as the impactors reacted with the atmosphere and consumed carbon dioxide that would have produced a greenhouse effect⁸³. Some suggest that elevated levels of greenhouse gases, most notably methane, prevented freezing⁸³. Methane sources may have included mantle degassing and impacting bodies^{83,85}. Others have proposed that solar ultraviolet radiation would cause rapid loss of methane and ammonia, and instead early Earth had a carbon dioxide- and hydrogen-rich atmosphere⁸⁸. Once methanogenic bacteria arose, the Earth may have been covered in a biogenic methane haze with surface temperatures higher than those we experience today⁸⁵. With the rise of oxygen levels roughly 2.3 Gyr ago and the corresponding loss of methane, the first global glaciation occurred^{84–86}. Biotic sequestration of carbon to form carbonate deposits and periodic release upon subduction may be the primary driver in the subsequent 'ice house'–'hot house' oscillations in climate^{89,90}. Hence, the emergence of life on Earth seems to have fundamentally altered the evolution of the planet.

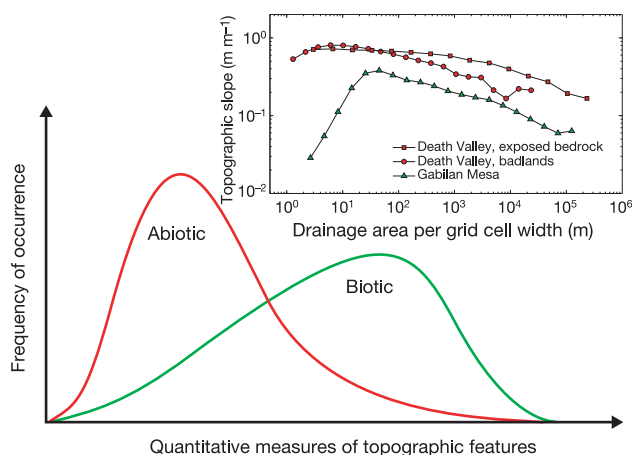


Figure 5 | Hypothetical frequency distributions of landform properties for the present Earth and an abiotic Earth. Whereas the same range of landform types would probably be present in both scenarios (provided the abiotic Earth retains its water, thereby permitting plate tectonics to occur), the frequency distributions of measurable landform properties (such as mountain height, steepness or curvature; the sinuosity of rivers; the extent of the landscape with a soil mantle; slope–area characteristics) might differ. Inset, relationships between slope (the magnitude of the topographic gradient) and drainage area per grid cell width for three landscapes. Biotic processes are of minimal erosional importance in areas of badlands and exposed bedrock in Death Valley, California, but are significant in the Gabilan Mesa, California, a landscape similar to that shown in Fig. 1. The differences between the slope–area relationships suggest several variables whose frequency distributions should differ in biotic and abiotic landscapes. Comparisons between landscapes with different spatial scales could also be made by normalizing the horizontal axis by topographic relief (the elevation difference between the highest and lowest points). Drainage area and slope were calculated from ALSM data with a horizontal resolution (grid cell width) of 2 m (provided by NCALM). Points are the mean slopes within logarithmically spaced, non-overlapping bins. Bars showing the standard errors of the binned values are smaller than the symbols. Areas sampled are 2.7 km² for Death Valley bedrock, 0.24 km² for Death Valley badlands, and 5.8 km² for the Gabilan Mesa.

become possible to define the frequency distributions of these attributes for Earth. Future missions to Mars will acquire data with sufficient resolution, and perhaps sufficient areal coverage, to permit a reasonably thorough survey of an abiotic landscape⁶³. Hence, we cannot at present compare the distributions of slope–area attributes or other landscape properties in biotic and abiotic landscapes, but it may become possible to do so in the near future.

There may be a unique topographic signature of life at the sub-metre scale. Burrowing organisms often create small mounds at the surface (for example, gopher mounds, worm castings, termite mounds and crab mounds). When trees topple, their root systems can pull up wads of soil, creating pit-and-mound topography. Constructional biotic features might have distinct forms and densities of occurrence, but until the resolution of topographic data becomes at least an order of magnitude finer than what is currently available, such features will remain virtually undetectable. The largest biotic constructional features on Earth, coral reefs, are effectively life itself rather than the influence of life on topography, and hence are not considered here.

Once biotic processes are explicitly included in geomorphic transport laws, landscape evolution models can be used to perform numerical experiments that quantitatively explore the effects of life's presence or absence on landscape form, in a manner analogous to modelling the climatic effects of a twofold increase in atmospheric CO₂. Initial experiments of this kind⁷, using current knowledge of geomorphic transport laws and approximations for the dynamics and erosional influence of vegetation, suggest that vegetation can affect landscape form by changing the dominant erosion process from overland flow under abiotic conditions to landsliding in well-vegetated states. While it may be difficult to test these predictions empirically on Earth, the ever-increasing volume of data for Mars should soon provide an opportunity for quantitative comparisons.

We currently lack most of the geomorphic transport laws required to formulate mechanistic landscape evolution models. Thus, the summary in Table 1 is both a tabulation of the current state of knowledge of geomorphic transport laws and, effectively, a list of future research directions. The role of biotic processes in existing laws is beginning to be quantified. Soil mass transport has been expressed in terms of the number density and energy expenditure of soil-moving organisms⁶⁴. Field and modelling studies of the effects of fire^{65,66} are revealing the role of vegetation in controlling the resisting forces in slope-dependent transport and shallow landsliding. Numerical experiments have begun to explore how vegetation influences erosion processes and landscape form^{6,7}. These early efforts point the way towards a mechanistic understanding of the role of biotic processes in landscape evolution and the possibility of identifying the topographic signature of life.

- Toy, T. J., Foster, G. R. & Renard, K. G. *Soil Erosion: Processes, Prediction, Measurement, and Control* (Wiley, New York, 2002).
- Rodríguez-Iturbe, I. & Rinaldo, A. *Fractal River Basins* (Cambridge Univ. Press, Cambridge, UK, 1997).
- Howard, A. D. in *Incised River Channels* (eds Darby, S. E. & Simons, A.) 277–299 (Wiley, New York, 1999).
- Roering, J. J., Almond, P., Tonkin, P. & McKean, J. Constraining climatic controls on hillslope dynamics using a coupled model for the transport of soil and tracers: Application to loess-mantled hillslopes, South Island, New Zealand. *J. Geophys. Res.* **109**, doi:10.1029/2003JF000034 (2004).
- Kirkby, M. J. Modelling the links between vegetation and landforms. *Geomorphology* **13**, 319–335 (1995).
- Collins, D. B. G., Bras, R. L. & Tucker, G. E. Modelling the effects of vegetation-erosion coupling on landscape evolution. *J. Geophys. Res.* **109**, doi:10.1029/2002JF000028 (2004).
- Istanbulluoglu, E. & Bras, R. L. Vegetation-modulated landscape evolution: Effects of vegetation on landscape processes, drainage density, and topography. *J. Geophys. Res.* **110**, F02012, doi:10.1029/2004JF000249 (2005).
- Willgoose, G., Bras, R. L. & Rodríguez-Iturbe, I. A coupled channel network growth and hillslope evolution model; 1, Theory. *Wat. Resour. Res.* **27**, 1671–1684 (1991).
- Howard, A. D. A detachment-limited model of drainage basin evolution. *Wat. Resour. Res.* **30**, 2261–2285 (1994).
- Tucker, G. E. & Bras, R. L. Hillslope processes, drainage density, and landscape morphology. *Wat. Resour. Res.* **34**, 2751–2764 (1998).
- Beaumont, C. P., Fullsack, P. & Hamilton, J. in *Thrust Tectonics* (ed. McClay, K. R.) 1–18 (Chapman and Hall, New York, 1992).
- Tucker, G. E. & Slingerland, R. L. Erosional dynamics, flexural isostasy, and long-lived escarpments: a numerical modeling study. *J. Geophys. Res.* **99**, 12229–12243 (1994).
- van der Beek, P. & Braun, J. Numerical modelling of landscape evolution on geological time-scales: A parameter analysis and comparison with the south-eastern highlands of Australia. *Basin Res.* **10**, 49–68 (1998).
- Willett, S. D. Orography and orogeny: The effects of erosion on the structure of mountain belts. *J. Geophys. Res.* **104**, 28957–28981 (1999).
- Tucker, G. E. & Slingerland, R. L. Drainage basin responses to climate change. *Wat. Resour. Res.* **33**, 2031–2047 (1997).
- Kirkby, M. J. in *Slopes: Form and Process* (ed. Brunsden, D.) 15–30 (Spec. Publ. 3, Institute of British Geographers, London, 1971).
- Dietrich, W. E. et al. in *Prediction in Geomorphology* (eds Wilcock, P. & Iverson, V.) 103–132 (AGU Geophys. Monogr. 135, American Geophysical Union, Washington DC, 2003).
- Smith, T. R. & Bretherton, F. P. Stability and the conservation of mass in drainage basin evolution. *Wat. Resour. Res.* **8**, 1506–1529 (1972).
- Whipple, K. X. Bedrock rivers and the geomorphology of active orogens. *Annu. Rev. Earth Planet. Sci.* **32**, 151–185 (2004).
- Gabet, E. M., Reichman, O. J. & Seabloom, E. W. The effects of bioturbation on soil processes and sediment transport. *Annu. Rev. Earth Planet. Sci.* **31**, 249–273 (2003).
- Heimsath, A. M., Dietrich, W. E., Nishiizumi, K. & Finkel, R. C. The soil production function and landscape equilibrium. *Nature* **388**, 358–361 (1997).
- Ehrlich, H. L. Geomicrobiology: its significance for geology. *Earth-Sci. Rev.* **45**, 45–60 (1998).
- Newman, D. K. & Banfield, J. F. Geomicrobiology: elucidation of the molecular-scale interactions that underpin biogeochemical systems. *Science* **296**, 1071–1077 (2002).
- Roering, J. J. Soil creep and convex-upward velocity profiles: theoretical and experimental investigation. *Earth Surf. Process. Landforms* **29**, 1597–1612 (2004).
- Fleming, R. W. & Johnson, A. M. Rates of seasonal creep of silty clay soil. *Q. J. Eng. Geol.* **8**, 1–29 (1975).
- Heimsath, A. M., Chappell, J., Spooner, N. A. & Questiaux, D. G. Creeping soil. *Geology* **30**, 111–114 (2002).
- Roering, J. J., Kirchner, J. W. & Dietrich, W. E. Evidence for non-linear, diffusive sediment transport on hillslopes and implications for landscape morphology. *Wat. Resour. Res.* **35**, 853–870 (1999).
- Furbish, D. J. & Fagherazzi, S. Stability of creeping soil and implications for hillslope evolution. *Wat. Resour. Res.* **37**, 2607–2618 (2001).
- Howard, A. D. & Kerby, G. Channel changes in badlands. *Geol. Soc. Am. Bull.* **94**, 739–752 (1983).
- Seidl, M. A. & Dietrich, W. E. The problem of channel erosion into bedrock. *Catena Suppl.* **23**, 101–124 (1992).
- Howard, A. D., Dietrich, W. E. & Seidl, M. A. Modeling fluvial erosion on regional to continental scales. *J. Geophys. Res.* **99**, 13971–13986 (1994).
- Sklar, L. & Dietrich, W. E. Sediment supply, grain size and rock strength controls on rates of river incision into bedrock. *Geology* **29**, 1087–1090 (2001).
- Sklar, L. & Dietrich, W. E. A mechanistic model for river incision into bedrock by saltating bedload. *Water Resour. Res.* **40**, W06301, doi:10.1029/2003WR002496 (2004).
- Riquelme, R., Martinod, J., Herail, G., Darrozes, J. & Charrier, R. A geomorphological approach to determining the Neogene to Recent tectonic deformation in the Coastal Cordillera of northern Chile (Atacama). *Tectonophysics* **361**, 255–275 (2003).
- Amundson, R. et al. Chemical alteration of soils on Earth as a function of precipitation: Insights into weathering processes relevant to Mars. *Eos* **85** (Fall Mtg Suppl.), abstr. P23B-06 (2004).
- Heimsath, A. M., Dietrich, W. E., Nishiizumi, K. & Finkel, R. C. Stochastic processes of soil production and transport: erosion rates, topographic variation, and cosmogenic nuclides in the Oregon Coast Range. *Earth Surf. Process. Landforms* **26**, 531–552 (2001).
- Heimsath, A. M., Furbish, D. & Dietrich, W. E. The illusion of diffusion: Field evidence for depth dependent sediment transport. *Geology* **33**, 949–952 (2005).
- Schumm, S. A. Patterns of alluvial rivers. *Annu. Rev. Earth Planet. Sci.* **13**, 5–27 (1985).
- Ward, P. D., Montgomery, D. R. & Smith, R. Altered river morphology in South Africa related to the Permian-Triassic extinction. *Science* **289**, 1740–1743 (2000).
- Micheli, E. R. & Kirchner, J. W. Effects of wet meadow riparian vegetation on streambank erosion. 2. Measurements of vegetated bank strength and consequences of failure mechanics. *Earth Surf. Process. Landforms* **27**, 687–697 (2003).
- Murray, A. B. & Paola, C. Modelling the effect of vegetation on channel pattern in bedload rivers. *Earth Surf. Process. Landforms* **28**, 131–143 (2003).
- Leopold, L. B., Wolman, M. G. & Miller, J. P. *Fluvial Processes in Geomorphology* (Freeman & Co., San Francisco, 1964).

43. Baker, V. C. *The Channels on Mars* (Univ. Texas Press, Austin, 1982).
44. Moore, J. M., Howard, A. D., Dietrich, W. E. & Schenk, P. M. Martian layered fluvial deposits: implications for Noachian climate scenarios. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL019002 (2003).
45. Malin, M. C. & Edgett, K. S. Evidence for persistent flow and aqueous sedimentation on early Mars. *Science* **302**, 1931–1934 (2003).
46. Bhattacharya, J. P., Payenberg, T. H. D., Lang, S. C. & Bourke, M. Dynamic river channels suggest a long-lived Noachian crater lake on Mars. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL022747 (2005).
47. Kleidon, A., Fraedrich, K. & Heimann, M. A green planet versus a desert world: Estimating the maximum effect of vegetation on the land surface climate. *Clim. Change* **44**, 471–493 (2000).
48. Hilley, G. E. & Strecker, M. R. Steady-state erosion of critical Coulomb wedges with applications to Taiwan and the Himalaya. *J. Geophys. Res.* **109**, B01411, doi:10.1029/2002JB002284 (2004).
49. Whipple, K. X. & Meade, B. J. Controls on the strength of coupling among climate, erosion, and deformation in two-sided, frictional orogenic wedges at steady state. *J. Geophys. Res.* **109**, F01011, doi:10.1029/2003JF000019 (2004).
50. Richards, M. A., Yang, W. S., Baumgardner, J. R. & Bunge, H. P. Role of a low-viscosity zone in stabilizing plate tectonics: Implications for comparative terrestrial planetology. *Geochem. Geophys. Geosyst.* **2**, doi:10.1029/2000GC000115 (2001).
51. Nimmo, F. & McKenzie, D. Volcanism and tectonics on Venus. *Annu. Rev. Earth Planet. Sci.* **26**, 23–51 (1998).
52. Chyba, C. F. Rethinking Earth's early atmosphere. *Science* **308**, 962–963 (2005).
53. Baker, V. R. Water and the martian landscape. *Nature* **412**, 228–236 (2001).
54. Craddock, R. A. & Howard, A. D. The case for rainfall on a warm, wet early Mars. *J. Geophys. Res.* **107**, doi:10.1029/2001JE001505 (2002).
55. Irwin, R. P., Craddock, R. A. & Howard, A. D. Interior channels in Martian valley networks: Discharge and runoff production. *Geology* **33**, 489–492 (2005).
56. Howard, A. D., Moore, J. M. & Irwin, R. P. III An intense terminal epoch of widespread fluvial activity on early Mars: 1. Valley network incision and associated deposits. *J. Geophys. Res.* **110**, doi:10.1029/2005JE002459 (2005).
57. Head, J. W. *et al.* Tropical to mid-latitude snow and ice accumulation, flow and glaciation on Mars. *Nature* **434**, 346–351 (2005).
58. Piqueux, S., Byrne, S. & Richardson, M. I. Sublimation of Mars's southern seasonal CO₂ ice cap and the formation of spiders. *J. Geophys. Res.* **108**, doi:10.1029/2002JE002007 (2003).
59. Byrne, S. & Ingersoll, A. P. A sublimation model for Martian South Polar ice features. *Science* **299**, 1051–1053 (2003).
60. Tomasko, M. G. *et al.* Rain, winds and haze during the Huygens probe's descent to Titan's surface. *Nature* **438**, 765–778 (2005).
61. Willgoose, G. A statistic for testing the elevation characteristics of landscape simulation models. *J. Geophys. Res.* **99**, 13987–13996 (1994).
62. Stock, J. D. & Dietrich, W. E. Valley incision by debris flows: evidence of a topographic signature. *Wat. Resour. Res.* **39**, 1089, doi:10.1029/2001WR001057 (2003).
63. McEwen, A. *et al.* MRO's High Resolution Imaging Science Experiment (HiRISE): Science expectations. *Sixth Int. Conf. on Mars* 3217 (2003); (<http://www.lpi.usra.edu/meetings/sixthmars2003/>) (2003).
64. Yoo, K., Amundson, R., Heimsath, A. M. & Dietrich, W. E. Process-based model linking pocket gopher (*Thomomys bottae*) activity to sediment transport and soil thickness. *Geology* **33**, 917–920 (2005).
65. Gabet, E. & Dunne, T. A stochastic sediment delivery model for a steep Mediterranean landscape. *Wat. Resour. Res.* **39**, 1237, doi:10.1029/2003WR002341 (2003).
66. Roering, J. J. & Gerber, M. Fire and the evolution of steep, soil-mantled landscapes. *Geology* **33**, 349–352 (2005).
67. Ahnert, F. in *L'evolution des Versants* (ed. Macar, P.) 23–41 (Univ. Liege, Liege, France, 1967).
68. Cox, N. J. On the relationship between bedrock lowering and regolith thickness. *Earth Surf. Process. Landforms* **5**, 271–274 (1980).
69. Anderson, R. S. Modeling the tor-dotted crests, bedrock edges, and parabolic profiles of high alpine surfaces of the Wind River Range, Wyoming. *Geomorphology* **46**, 35–58 (2002).
70. Culling, W. E. H. Analytical theory of erosion. *J. Geol.* **68**, 336–344 (1960).
71. Perron, J. T., Dietrich, W. E., Howard, A. D., McKean, J. A. & Pettinga, J. Ice-driven creep on Martian debris slopes. *Geophys. Res. Lett.* **30**, 1747, doi:10.1029/2003GL017603 (2003).
72. Iverson, R. M. Unsteady, nonuniform landslide motion, 2. Linearized theory and the kinematics of transient-response. *J. Geol.* **94**, 349–364 (1986).
73. Densmore, A. L., Ellis, M. A. & Anderson, R. S. Landsliding and the evolution of normal-fault-bounded mountains. *J. Geophys. Res.* **103**, 15203–15219 (1998).
74. Iverson, R. M. Regulation of landslide motion by dilatancy and pore pressure feedback. *J. Geophys. Res.* **110**, doi:10.1029/2004JF000268 (2005).
75. Stock, J. D. *Incision of Steepland Valleys by Debris Flows* Ph.D. thesis, Univ. California, Berkeley (2003).
76. MacGregor, K. R., Anderson, R. S., Anderson, S. P. & Waddington, E. D. Numerical simulations of glacial-valley longitudinal profile evolution. *Geology* **28**, 1031–1034 (2000).
77. Hildes, D. H. D., Clarke, G. K. C., Flowers, G. E. & Marshall, S. J. Subglacial erosion and englacial sediment transport modelled for North American ice sheets. *Quat. Sci. Rev.* **23**, 409–430 (2004).
78. Bagnold, R. A. *The Physics of Blown Sand and Desert Dunes* (Methuen, London, 1941).
79. Werner, B. T. A steady state model of wind-blown sand transport. *J. Geol.* **98**, 1–17 (1990).
80. McEwan, I. K. *The Physics of Sand Transport by Wind* Ph.D. thesis, Univ. Aberdeen (1991).
81. Anderson, R. S. Erosion profiles due to particles entrained by wind: Application of an eolian sediment-transport model. *Geol. Soc. Am. Bull.* **97**, 1270–1278 (1986).
82. Lancaster, N. & Baas, A. Influence of vegetation cover on sand transport by wind: field studies at Owens Lake, California. *Earth Surf. Process. Landforms* **23**, 69–82 (1998).
83. Sleep, N. H. & Zahnle, K. Carbon dioxide cycling and implications for climate on ancient Earth. *J. Geophys. Res.* **106**, 1373–1400 (2001).
84. Kump, L. R., Brantley, S. L. & Arthur, M. A. Chemical weathering, atmospheric CO₂, and climate. *Annu. Rev. Earth Planet. Sci.* **28**, 611–667 (2000).
85. Kasting, J. F. & Catling, D. Evolution of a habitable planet. *Annu. Rev. Astron. Astrophys.* **41**, 429–463 (2003).
86. Brasier, M., Green, O., Lindsay, J. & Steele, A. Earth's oldest (~3.5 Ga) fossils and the 'Early Eden hypothesis': questioning the evidence. *Orig. Life Evol. Biosph.* **34**, 257–269 (2004).
87. Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).
88. Tian, F., Toon, O. B., Pavlov, A. A. & De Sterck, H. A hydrogen-rich early Earth atmosphere. *Science* **308**, 1014–1017 (2005).
89. Schrag, D. P., Berner, R. A., Hoffman, P. F. & Halverson, G. P. On the initiation of snowball Earth. *Geochem. Geophys. Geosyst.* **3**, doi:10.1029/2001GC00219 (2002).
90. Edmond, J. M. & Huh, Y. Non-steady carbonate recycling and implications for the evolution of atmospheric PCO₂. *Earth Planet. Sci. Lett.* **216**, 125–139 (2003).

Acknowledgements Many colleagues offered advice and direction in preparation of this review, especially I. Fung, D. Schrag, N. Sleep, L. Sklar and L. Kump. A. Howard, D. Furbish, M. Power and G. Hilley gave comments on earlier drafts of the paper. M. Gabet made several suggestions. A. Kleidon shared unpublished data. N. Snyder and the NSF National Center for Airborne Laser Mapping provided the topographic data for Death Valley. Aspects of this work were supported by the NSF National Center for Earth Surface Dynamics, the NSF Graduate Research Fellowship Program, and NASA (for work in the Atacama Desert with R. Amundson).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to W.E.D. (bill@eps.berkeley.edu).

Atomic packing and short-to-medium-range order in metallic glasses

H. W. Sheng¹, W. K. Luo¹, F. M. Alamgir², J. M. Bai³ & E. Ma¹

Unlike the well-defined long-range order that characterizes crystalline metals, the atomic arrangements in amorphous alloys remain mysterious at present. Despite intense research activity on metallic glasses and relentless pursuit of their structural description, the details of how the atoms are packed in amorphous metals are generally far less understood than for the case of network-forming glasses. Here we use a combination of state-of-the-art experimental and computational techniques to resolve the atomic-level structure of amorphous alloys. By analysing a range of model binary systems that involve different chemistry and atomic size ratios, we elucidate the different types of short-range order as well as the nature of the medium-range order. Our findings provide a reality check for the atomic structural models proposed over the years, and have implications for understanding the nature, forming ability and properties of metallic glasses.

Historically, a widely cited structural model for metallic glasses (MGs) is that of Bernal's dense random packing of hard spheres^{1,2}. It is now understood that Bernal's idea can satisfactorily model monatomic systems (or alloys with constituent species having comparable atomic sizes and insignificant chemical short-range order, SRO)³, but fail to describe many binary MGs, notably metal-metalloid glasses, where the chemical SRO is pronounced. The stereochemically defined model proposed later^{4,5}, in contrast, stipulates that the local unit (such as nearest neighbours) in amorphous alloy has the same type of structure as their crystalline compounds with similar composition. The general applicability of this model is still being debated, as experimental evidence has not yet been conclusive. Even less understood is the medium-range order (MRO), which can be defined as the next-level structural organization beyond the SRO⁶, for example, how the local structural 'units' are connected and arranged to fill three-dimensional (3D) space. The characteristics of the MRO remain one of the most important outstanding questions^{7,8} in MG research. Taking the earlier models further, a recent model⁹ proposes face-centred cubic (f.c.c.) packing of overlapping clusters as the building scheme for MG structures. The underlying principle is the efficient filling of space^{9,10}. Clearly, there is a pressing need for an in-depth reality check of these previous structural concepts, and for establishing a realistic structural picture of MGs.

Obtaining the 3D structure of metallic glasses

We set out to solve the 3D structure of MGs rather than proposing models *a priori*. To achieve a realistic description of the structure, we used experimental and computational techniques that allow us to explicitly consider chemical effects and atomic relaxations in realistic many-body interactions of 'soft atoms' (rather than hard spheres), which are absent in earlier models. Atomic coordinates from the 3D structure were then used for more detailed analysis of structural features, from which basic principles underlying the short-to-medium-range order could be extracted. As illustrated in detail below, our results validate several of the earlier concepts regarding the SRO and efficient packing, specify the particular topologies of cluster-like

entities, and suggest a fundamentally new cluster packing scheme that constitutes the MRO. For completeness, we also illustrate how the solute atoms arrange themselves in alloys with enriched solute concentrations.

Metallic glass structure from XRD, EXAFS and RMC. We started by analysing a model MG, Ni₈₀P₂₀, using structural information acquired through synchrotron X-ray measurements. In addition to wide-angle X-ray scattering (Fig. 1a), extended X-ray absorption fine structure (EXAFS) experiments have been used as the species-specific probe to provide local coordination information (see Fig. 1 and the Methods). Unlike previous work that dealt mostly with the one-dimensional pair distribution functions, we focused on the 3D atomic configuration (atomic positions, density) reconstructed using the reverse Monte Carlo (RMC)^{11,12} method, which produced the excellent fits to the experimental spectra in Fig. 1 (see Methods). The final configuration was double-checked by comparing the X-ray absorption near-edge structure (XANES) patterns calculated from the RMC structure (Fig. 1d) with that observed in the X-ray absorption experiments (shown in Supplementary Fig. S1 but not used in RMC fitting). We also confirmed that the one-dimensional partial pair distribution functions obtained from the RMC configuration are in satisfactory agreement with those obtained before from neutron isotope experimental measurements¹³ (see the comparison in Supplementary Fig. S2).

Structure from *ab initio* molecular dynamics simulations. We also pursued the 3D MG structure via a completely independent route: *ab initio* molecular dynamics simulations^{14,15}. Here the atomic forces are determined on the basis of first-principles calculations without the need for any experimental data or empirical potentials (details in the Methods^{16–21}). The *ab initio* calculations allowed us not only to obtain and analyse the static amorphous structure, but also to monitor the structural evolution during the liquid cooling process. More importantly, the *ab initio* modelling enabled us to deduce a general trend by systematically comparing a number of MG systems with different chemical make-ups (such as transition metal (TM)–TM and TM–metalloid), different atomic size ratios and different levels of solute concentrations.

¹Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA. ²National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA. ³High-Temperature Materials Laboratory, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.

Figure 2 illustrates the evolution of the reduced radial distribution functions (RRDFs) as a function of temperature during the cooling of the *ab initio* Ni–P system. The development of SRO with decreasing temperature is clearly observed from the structure developed in the second peak. The partial RRDFs of the final 300 K configuration (purple curve) agree remarkably well with those directly measured from experiments¹³ (orange dots). Furthermore, the EXAFS spectra in Fig. 1 (as well as the XANES spectra in Supplementary Fig. S1) obtained using first-principles calculations²² for the 300 K *ab initio* configurations also exhibit an excellent match with the experimental data. These validations lend strong support to the use of the

configuration obtained from the ‘virtual’ quenching experiments for structural analysis of the 3D atomic packing. An additional check for both the RMC and molecular dynamics configurations is that their mass (or atomic) densities are all within 1 to 2% of the experimentally measured value for the Ni–P (and 0.7% for Ni–B discussed below) MG.

Structural analysis based on Voronoi tessellation

The atomic positions in both the RMC and *ab initio* molecular dynamics configurations can then be used, and compared, to unravel the short-to-medium-range details. Our first objective was to identify the possible presence of building blocks, or ‘clusters’. The first step was to resolve the local nearest-neighbour coordination. The coordination number (CN) can be determined unambiguously by using the Voronoi tessellation method^{23–26}, which also characterizes the local atomic environment. We use the Voronoi index $\langle n_3, n_4, n_5, n_6, \dots \rangle$, where n_i denotes the number of i -edged faces of the Voronoi polyhedron and $\sum_i n_i$ is the total CN, to designate and differentiate the type of the coordination polyhedron surrounding a centre solute atom. For example, tri-capped trigonal prism packing (TTP) corresponds to a Voronoi index of $\langle 0, 3, 6, 0 \rangle$, while a mono-capped square archimedean antiprism (slightly distorted from the TTP) has a Voronoi index of $\langle 0, 5, 4, 0 \rangle$ (see Supplementary Fig. S3).

In the Ni–P glass the solute P was found to be coordinated in the first neighbouring shell by Ni atoms only, and the statistics of the Voronoi analysis indicated that the average first-neighbour CN of the P solute is 10.5 for the *ab initio* configuration, and 10.6 for the RMC model, with very similar distribution for the two 3D configurations (see the distribution in Fig. 3 for the *ab initio* configuration). Such a CN would seem to be larger than was determined in previous

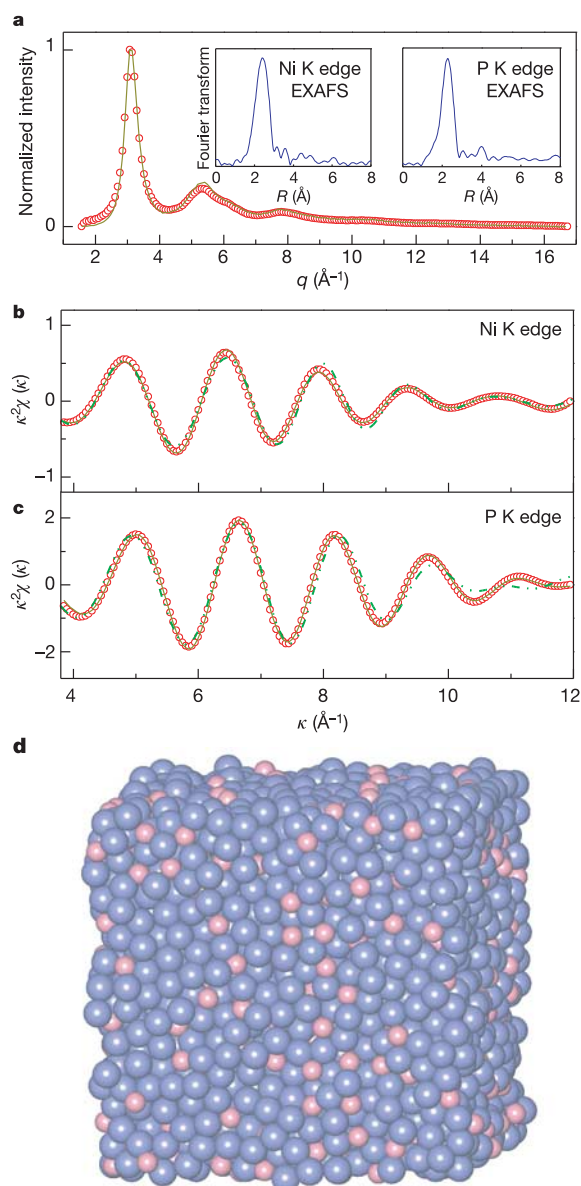


Figure 1 | Reverse Monte Carlo modelling to reproduce the experimental X-ray diffraction and absorption data. **a**, Solid lines are the experimental XRD and Fourier-transformed EXAFS spectra for the Ni₈₀P₂₀ amorphous alloy. The circles are for the XRD pattern calculated for the eventual RMC configuration. **b**, **c**, The solid lines present the inverse Fourier transforms of the first peaks in the Fourier-transformed EXAFS data in **a**. $\chi(\kappa)$ is the EXAFS, where κ is the photoelectron wavevector. R is the distance between atoms. The EXAFS spectra calculated for the RMC configuration (circles) and for the *ab initio* molecular dynamics configuration (green line) are in good agreement with the experimental data. **d**, A view of the final RMC configuration. The blue and pink balls represent Ni and P atoms, respectively.

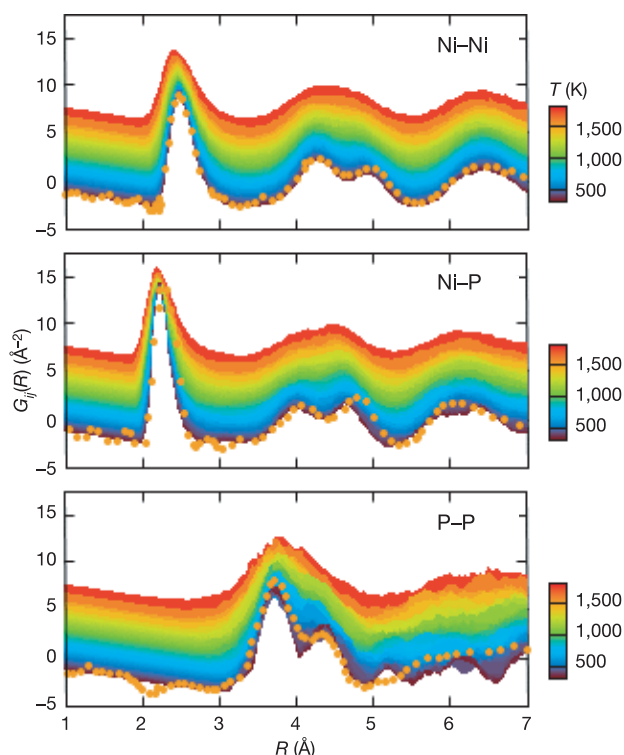


Figure 2 | The evolution of partial RRDFs of the Ni₈₀P₂₀ liquid observed in *ab initio* molecular dynamics simulations. The partial RRDFs, $G_{ij}(R)$, change as a function of temperature during the cooling and glass formation of the Ni₈₀P₂₀ liquid. $G_{ij}(R) = 4\pi R[\rho_{ij}(R)/c_j - \rho_0]$, where $\rho_{ij}(R)$ is the atomic density of the j atom at distance R from an i atom, and ρ_0 is the mean atomic density. The experimental data for this glass are also shown for comparison (orange circles, redrawn from ref. 13).

RDF work^{7,13} by counting the number of atoms belonging to the 'first shell', which is very sensitive to the predefined cut-off distance. However, we confirmed that the results are consistent, as the CN turned out to be 9.4 when we used a cut-off distance of 2.8 Å in the RDF (Fig. 2), in good agreement with previous analyses of TM-metalloid MGs^{7,13}.

Dependence of local packing on atomic size ratio

To observe general features and trends, we analysed a number of model MGs using *ab initio* molecular dynamics simulations. These structural configurations are validated by comparing the calculated RRDFs with the experimentally measured ones (see below). Owing to the different atomic size ratios, the CN distribution of these MGs shifts consistently, as shown in Fig. 3. The solvent atoms surrounding the centre solute atom apparently form different types of coordination polyhedra in these various systems. As opposed to Pauling's rule of parsimony²⁷, which states that the number of the chemically different environments of an (ionic) compound tends to be small (for example, even the most complex of silicates minerals typically has only three to four different kinds of coordination polyhedra²⁸), the solute-centred polyhedra in each of the MGs studied involve a large number (as many as a few hundred) of coordination polyhedron types. Such diversity has previously been reported in classical molecular dynamics simulations of model systems^{23,29}, and has recently been enumerated^{25,26}. This finding suggests that the local arrangements cannot be modelled by a uniquely prescribed stereochemical structure. This conclusion was also corroborated by our analysis of the rather broad bond-angle distribution (see Supplementary Fig. S4).

On the other hand, despite the large number of the geometrically distinct polyhedron types, certain polyhedra appeared with high frequencies (20–40%; see Fig. 4). A closer inspection suggests that they are Kasper polyhedra, which are the deltahedra that involve the minimum number of disclinations^{30,31}. The atomic packing configurations of different Kasper polyhedra are shown in Fig. 3. In the Ni–P glass, the percentages of the Z10 ($Z = \text{CN}$) Kasper polyhedron (the bi-capped square antiprism, BSAP, with Voronoi index $\langle 0,2,8,0 \rangle$), and the Z11 Kasper polyhedra (with $\langle 0,2,8,1 \rangle$) are

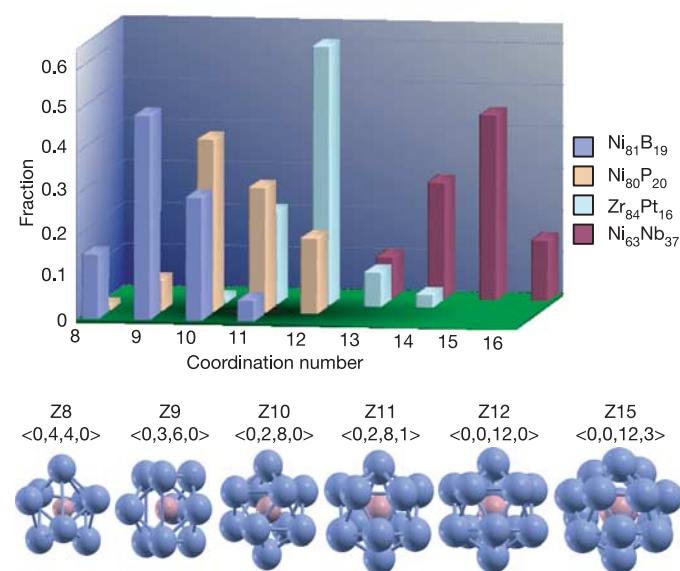


Figure 3 | CN distribution of the solute atoms in several representative MGs, obtained from *ab initio* calculations. The average CN changes with the effective atomic size ratio, and for each glass the majority of the solute atoms (>75% of total) have two dominant CNs. Also shown are the Kasper polyhedra corresponding to the different CNs. The Kasper polyhedra are the dominant coordination polyhedra in the relaxed MGs.

16.2% and 10.6%, respectively, while in the Ni₈₁B₁₉ MG, the Z9 Kasper polyhedron (TTP, with $\langle 0,3,6,0 \rangle$) and the Z10 Kasper polyhedra are dominant (at 17.8% and 7.1%, respectively). In comparison, prominent icosahedral local order is found in the Zr₈₄Pt₁₆ MG, where the occurrences of the Z12 and Z11 Kasper polyhedra are 14.1% and 9.0%, respectively. Therefore, it is evident that the Kasper polyhedron SRO is the main underlying topological SRO in the MGs.

The preference for a particular type, as seen in Fig. 3 together with the shift of the main peak, is controlled by the effective atomic size ratio between the solute and solvent atoms, R^* . With decreasing R^* , the preferred polyhedra type changes from the Frank–Kasper³² type (for $R^* > 1.2$) to the icosahedral type ($R^* \approx 0.902$, as for Zr–Pt with $R^* = 0.90$), and then to the BSAP type ($R^* \approx 0.835$, as in Ni–P with $R^* = 0.78$), and then to the TTP type ($R^* \approx 0.732$, as in Ni–B with $R^* = 0.69$); see Figs 3 and 4. As pointed out above, some MGs, notably TM-metalloid systems (Ni–B, Pd–Si and so on), indeed exhibit the TTP local order, similar to that in corresponding crystalline compounds³³. But this should be understood as a geometrical consequence and only specific to a small number of MGs that meet the R^* requirement for the formation of TTP ($R^* \approx 0.732$). Our results not only validate the important role of the relative size

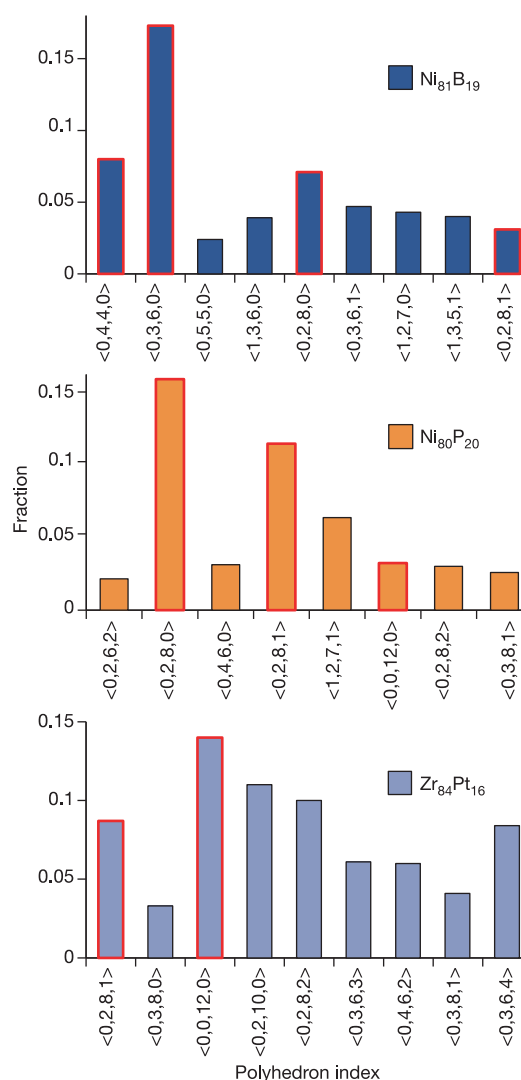


Figure 4 | The occurrences of different coordination polyhedra (with different Voronoi indices) of the solute atoms in the MGs. The bars outlined in red show the frequency of the dominant Kasper polyhedra. Note that only the polyhedra with a population of >2% are shown.

between the solute and solvent atoms^{9,10,34}, but also explicitly specify the packing topologies for the various CNs.

Packing of quasi-equivalent clusters

For each one of the MGs, the several types of local coordination polyhedra specified above are geometrically different, not identical in topology and CN. They can nevertheless be considered quasi-equivalent, or cluster-like units for a given glass, supporting the framework of cluster packing⁹. This is because, first, the solute-centred 'clusters' in Fig. 5 (dashed circles) have similar sizes, as seen from their narrow volume distribution. Second, as discussed above, the CN of the solute at centre has a relatively small variance around an average controlled by R^* , and the polyhedra can be considered distortions of certain specific types of Kasper polyhedra. That the CN is not a unique integer for a given glass, as is sometimes designated in idealized hard-sphere models^{9,10}, is not surprising. This moderate CN distribution (or a range of quasi-equivalent clusters) is a natural consequence of strain relaxation in an MG to allow for more comfortable packing of 'soft' atoms (many-body interactions rather than touching hard spheres all of the same size), in the entire solid (not just one nearest-neighbouring shell as idealized in some models¹⁰). Such arrangements will permit more flexibility for efficient filling of the entire 3D space, with reduced energy (see below).

We point out that the formation of the local solute-centred coordination polyhedra is a manifestation of the strong chemical SRO favouring unlike bonds. An examination of the calculated electronic structures demonstrates the strong chemical affinity between the TM and metalloid, resulting from the partially covalent nature of the bonding that is still mostly non-directional. In the TM–TM glasses, we often observed non-additive pair interaction, arising from the charge transfer and screening of d electrons. Chemical SRO is well known in these TM–metalloid and TM–TM systems, and is usually one of the prerequisites for the easy formation of MGs³⁵.

Owing to the bond non-additivity, the effective R^* is different from that estimated from their Goldschmidt atomic radii. In the literature, the absence of the direct solute–solute contacts is termed 'solute–solute avoidance'⁷. In other words, in the partial RRDFs of the solutes (see TM–metalloid glasses in Fig. 2 and Supplementary Fig. S7), there is little or no solute–solute pair correlation peak at the nearest-neighbour position. The chemical SRO—that is, the fact that the solute atom sits in the centre defining the quasi-equivalent clusters—sets the stage for the formation of the type of MRO discussed below.

The solute–solute correlations and MRO. With the 'quasi-equivalent cluster' defined in this way, and after illustrating the intra-cluster packing (topological and chemical SRO), the next issue we discuss is the correlation among solute atoms (and solute-centred clusters). This addresses the important question of how the clusters are connected and packed to fill the 3D space, giving rise to MRO.

We now describe the (dense) geometric/topological packing of the quasi-equivalent clusters. To differentiate several important candidate schemes, we make use of the common-neighbour analysis³⁶. In the present study, the quasi-equivalent clusters were treated as rigid balls, each occupying the volume of the coordination polyhedron of the solute atom at centre. As shown in Fig. 5a, in the $\text{Ni}_{81}\text{B}_{19}$, $\text{Ni}_{80}\text{P}_{20}$ and $\text{Zr}_{84}\text{Pt}_{16}$ systems, the solute-centred clusters are packed with appreciable icosahedral topological order, as evidenced from the prominent presence of the 555 icosahedral signature pairs, as well as the 544 and 433 pairs (the 555 index represents a perfect fivefold ring of common neighbours and the latter two correspond to the defective/distorted rings³⁶). Very similar results were found in the RMC configuration of the Ni–P system; see Supplementary Fig. S5. The 421 and 422 indices, signatures for close-packed face-centred cubic (f.c.c.) and hexagonal close-packed (h.c.p.), are insignificant.

Note that the common-neighbour analysis indices above reveal

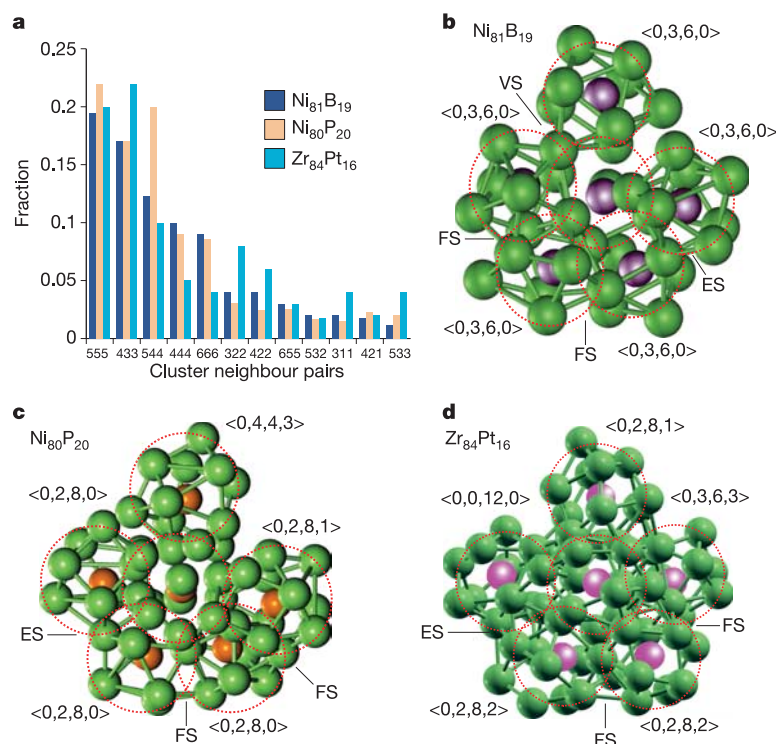


Figure 5 | The packing of the solute-centred quasi-equivalent clusters, showing their MRO. **a**, The cluster common-neighbour analysis showing that the local clusters in the MGs exhibit icosahedral type ordering. The typical cluster connections, exhibiting the fivefold symmetry, are detailed for $\text{Ni}_{81}\text{B}_{19}$, $\text{Ni}_{80}\text{P}_{20}$ and $\text{Zr}_{84}\text{Pt}_{16}$ in **b**, **c** and **d**, respectively. For each local

cluster, the Voronoi index is given to indicate its identity. FS, ES and VS denote face-sharing, edge-sharing and vertex-sharing, respectively. The red dashed circles delineate the clusters and suggest the quasi-equivalence of the clusters. The bonds for the centre cluster are not connected for clarity.

the presence of fragments (see Supplementary Fig. S5) of icosahedra, but not necessarily any perfect icosahedra. This can be termed 'icosahedral ordering', a term often used to describe local topological SRO in undercooled liquids^{37,38} and glasses¹² (such as the Zr–Pt glass discussed here). Here it is short range for the packing of clusters, but already medium range from the standpoint of atomic (solute) correlation beyond one cluster. The icosahedral-type MRO (cluster) ordering is a general feature regardless of the types of SRO inside the clusters, as shown in Fig. 5b–d, where the intracluster packing varies from the TTP order in Ni–B, to the BSAP in Ni–P, to the icosahedral SRO in Zr–Pt, as illustrated above. The tendency for icosahedral order is not surprising, as in all these alloy systems, each of the clusters has an average of 12–13 neighbouring clusters of about the same size (see Supplementary Fig. S6 for the cluster–cluster common-neighbour analysis). This indicates dense packing, hence favouring f.c.c., h.c.p., or icosahedral order. The icosahedral packing scheme preferring fivefold symmetry is in fact intrinsically associated with the dense packing in amorphous materials, and is not uncommon in the literature for the packing of small clusters or particles^{36,39–41}.

We stress that the *ab initio* calculated structures produce partial RRDFs in very good agreement with the experimental ones: see the excellent match of the RRDFs, Supplementary Fig. S7. Note also in Supplementary Fig. S7 that the splitting B–B first correlation peak seen in the experimental RRDF of Ni–B (and solute–solute correlation in some other TM–metalloid systems as well) was not reproduced by the previously proposed model assuming f.c.c. packing of overlapping equal-sized clusters⁹. This strongly suggests that the icosahedral fivefold packing is a more realistic ordering pattern of cluster–cluster connection in MGs. Our solute–solute–solute bond-angle analysis corroborates this conclusion. The cluster-connection diagrams in Fig. 5b–d for the several MGs illustrate the specific packing/connection schemes of the quasi-equivalent clusters, through the sharing of edges, faces and vertices. In fact, these different sharing schemes explain the different solute–solute distances that lead to the splitting in the peaks in the respective solute–solute partial RRDFs (Supplementary Fig. S7). Similar schemes for the communication among clusters were predicted in Gaskell's

model for the connection of TTP clusters⁷, but in a random chain configuration.

Cavities. As suggested in the previously proposed model of f.c.c. packing of stoichiometric clusters⁹, the packing of quasi-equivalent clusters may leave behind some 'cavities'. They can be identified by locating spaces with relatively low electron charge densities. We found that the larger cavities are associated with multiple clusters, and have a correlation with the cluster arrangement (See Supplementary Fig. S8). The location of the larger cavities resembles Bernal's canonical holes², if regarding the clusters as hard spheres. The analysis of the size and distribution of the cavities would have implications for glass-forming ability, especially when additional solute species with different sizes are introduced to fill such cavities to form bulk MGs⁴², as well as in the understanding of transport mechanism, glass transition, shear transformation in plastic flow, and the structure factors in scattering experiments (for example, the pre-peaks observed in some systems). But this topic is beyond the scope of this paper.

Strings and networks

The icosahedral ordering of single-solute-centred quasi-equivalent clusters is an efficient packing scheme, but is not the only type of MRO. When the concentration of the solute species increases to beyond the level at which each of all the solute atoms can be engaged in a single-solute-centred cluster, neighbour contact between solute atoms becomes inevitable. This produces direct bonding between like atoms. It is tempting to surmise that the excess solute atoms might be located at the interstitial 'vacancies' of the densely packed clusters⁹. But we found instead that a different kind of building unit emerges. In our *ab initio* calculated systems, the connection between solute atoms takes the form of 'strings'. For instance, two neighbouring solute atoms form an isolated pair surrounded by the solvent atoms, and the resulting configuration may be termed 'extended clusters'. In the $Zr_{80}Pt_{20}$ metallic glass (Fig. 6a) a solute pair is seen to have a total CN of ~ 17 . Similarly, string-like solute atoms are apparent in systems with higher solute concentrations (for example, Fig. 6b for amorphous $Al_{75}Ni_{25}$). The dense packing of the extended clusters, via edge-, vertex-, or face-sharing, constitutes the second type of MRO. This type of packing becomes significant when the solute concentration reaches roughly 20–30%, depending on the R^* ratio.

As the number of the direct bonds among solutes further increases with enriched solute concentration, more and more solute–solute contacts become unavoidable. A pronounced first-neighbour solute–solute pair correlation peak in the partial RRDFs emerges, as measured before for glasses such as Ni–Nb (ref. 43; Supplementary Fig. S7). When the solute interconnection percolation threshold is eventually reached, a network of the solute atoms takes form (Fig. 6c for amorphous $Ni_{63}Nb_{37}$). The string- or ring-like interconnection reduces the number of like bonds (increases the number of unlike bonds), when compared to aggregated solute atoms, leading to energy reduction. This is illustrated in Supplementary Fig. S9: when the $Zr_{70}Pd_{30}$ glass⁴⁴ is cooled from 1,800 to 300 K, the solute–solute connection becomes more network-like. The acute (solute–solute–solute) bond angles observed at high temperatures disappear, reflecting the tendency to order the solute atoms in string-like and ring-like medium-range arrangements that reduce the number of like bonds. The solute–solute CN in this open and extended connection is generally no more than 3 to 7. This network-type arrangement of the solute atoms leads to a much-involved third type of MRO. The concept of connected random network has previously been proposed in ref. 45. We have thus uncovered a whole spectrum of atomic packing schemes, which vary not only with R^* but also with solute concentration.

Limitations of models

We should briefly discuss the inherent limitations of the modelling techniques. The rapid quenching used for the *ab initio* systems did

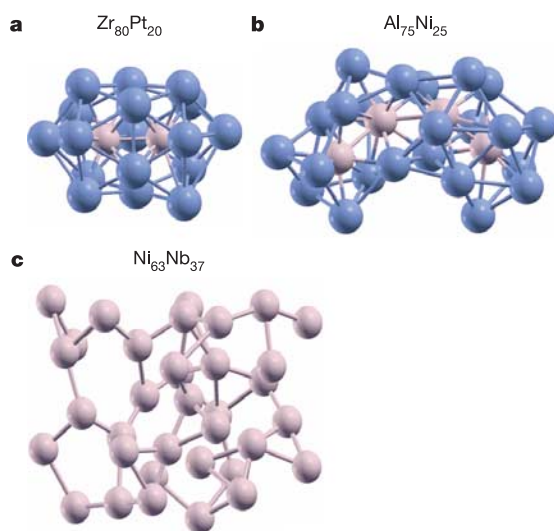


Figure 6 | Configurations of solute atoms at increasing solute concentrations. **a**, Extended clusters in the form of 'lone pairs' (here, two atom-sharing icosahedra with CN = 17) in the $Zr_{80}Pt_{20}$ MG. **b**, Extended clusters in the form of 'strings' in the $Al_{75}Ni_{25}$ MG. The dense packing of such 'extended' clusters constitutes the second type of MRO. **c**, The solute network formed in the $Ni_{63}Nb_{37}$ MG, in which the solute concentration is high. The network-like arrangement of the solute atoms gives rise to a different, third type of MRO.

not appear to suppress the development of ordering, as the agreement between the calculated (at ultrahigh quenching rates) and experimental (for glasses obtained at lower cooling rates) partial RRDs is remarkable (Fig. 2 and Supplementary Fig. S7). This shows that the pair correlations established in our calculated systems are similar to those in experimental samples, for the SRO and the MRO on the short-range side⁴⁶ examined here (the experimental data available for validation are for the range below 8 Å, although in general MRO may extend up to approximately 2 nm; see Fig. 2 and Supplementary Fig. S7). We also compared the structure of the Ni₈₀P₂₀ MG obtained at two different quench rates (see Methods) and found no obvious difference.

To observe the effects of the small size of the system and the imposed periodic boundary conditions, we examined different sizes up to 2.2 nm for the edge length of the cube, and also modelled the Ni₈₀P₂₀ glass using a large system of 4,000 atoms. In the latter case, however, classical molecular dynamics simulations¹⁶ involving empirical or hypothetical interatomic potentials had to be used (binary mixture of Lennard-Jones particles using the modified Kob–Andersen interatomic potentials⁴⁷). All the results are consistent with what we have presented above. This suggests the particular MRO is not due to a small system under periodic boundary conditions. However, it is certainly true that one prefers slower cooling rates to allow the system to survey the energy landscape more thoroughly, attaining lower energy states through increased structural ordering. While SRO may converge fast, the degree and extent of the MRO development may be affected by the rapid computer quench in small systems (our RMC system is much larger, but the Monte Carlo procedure used in RMC also favours randomization). Therefore, the SRO and MRO we report here could be underestimates; the ordering actually present in the (more structurally relaxed) MGs should be at least as much.

METHODS

EXAFS data. Ni₈₀P₂₀ foils with thicknesses of up to 10 µm were prepared using electroplating from Brenner-type baths onto Cu foils, which were etched away after deposition. X-ray diffraction (XRD), transmission-mode Ni K-edge and fluorescence-mode P K-edge EXAFS data were collected at beamline X14A and X15B, respectively, of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL). The EXAFS data were Fourier-transformed into real space after pre-edge and post-edge background subtractions. The first peak up to ~3.5 Å in real space corresponding to the nearest-neighbour shell was inverse-Fourier-transformed back into *k* space. Through the Fourier-filtering procedure, the simple single-scattering signal from local environment around the absorbing element was decomposed from the total EXAFS signal and can be analysed separately using standard EXAFS formula and theoretical amplitude and phase functions.

RMC modelling. RMC modelling, based on experimental XRD and Fourier-filtered EXAFS data above, was employed to reconstruct 3D atomic configuration in the amorphous alloy without any predetermined hypothetical structural models. The RMC system used typically 4,000 atoms. The metropolis algorithm is used in the RMC code and the error between the simulated and experimental data are analogous to the 'potential energy' in conventional Monte Carlo simulations. An atomic guess configuration is generated by randomly moving or exchanging atoms. The structure is improved by reducing the fitting residual while the deviation serves as 'temperature' and is slowly 'cooled' following the simulated annealing scheme. We confirmed that the output structure of our RMC modelling, upon matching the experimental input after ~10⁷ running steps, is independent of the starting configuration: the initial configuration can be either a 3D lattice or a random structure under appropriate constraints and periodic boundary conditions.

Ab initio simulations. For the *ab initio* molecular dynamics simulations, canonical NVT (constant number, volume, temperature) ensembles are used applying the Vienna *ab initio* simulation package¹⁷. The temperature was controlled using the Nose-Hoover thermostat¹⁸. 100–200 atoms with the desirable composition, are arranged in a cubic box with periodic boundary conditions. The simulation was performed on the *Γ* point only. Projected augmented planewaves (PAW)^{18,19} with the Perdew–Wang exchange–correlation potentials²⁰ have been adopted. The valence state of each element has been previously defined in the provided PAW pseudopotentials. One important

quantity throughout the simulations is the determination of the density of the systems. For this purpose, the ensemble was melted and equilibrated at high temperatures (1,800–2,200 K) for 2,000 timesteps, with each timestep of 5 fs. The conjugated gradient method was used to obtain the inherent structure²¹ of the liquid. The density was adjusted to correspond to the zero pressure of the inherent structure of the liquid. The quenching rate used to obtain the various glass systems is in the range of 200 to 20 K per 1,000 timesteps (4×10^{13} to 4×10^{12} K s⁻¹). Alloys selected include several TM–metalloid (Ni₈₀P₂₀, Ni₈₁B₁₉), TM–TM (Zr₈₄Pt₁₆, Zr₈₀Pt₂₀, Zr₇₀Pd₃₀, Ni₆₃Nb₃₇) and metal–TM (Al₇₅Ni₂₅, Al₂₅Zr₇₅) systems.

Voronoi tessellation. Details about the Voronoi tessellation method can be found in references²³. In constructing the Voronoi polyhedra, we removed those surfaces with area less than 1% of the total area of the polyhedron surfaces. By doing so, the degeneracy problem and the effects of thermal vibration are minimized. For the cluster–cluster common neighbour analysis, a pre-set threshold length for the nearest neighbour distance would introduce an ambiguity. We used instead the Voronoi tessellation method to determine the cluster neighbours. This was done by tessellating the space into 'regions of the solute atoms', and two neighbouring solute atoms represent two connecting clusters. The common-neighbour analysis was then carried out on the cluster neighbours identified. In the bond-angle analysis, the bond angle refers to the included angle between two atoms in the first coordination shell and the atom in the centre of the cluster.

Received 1 August; accepted 2 November 2005.

- Bernal, J. D. Geometry of the structure of monatomic liquids. *Nature* **185**, 68–70 (1960).
- Bernal, J. D. The structure of liquids. *Proc. R. Soc. Lond. A* **280**, 299–322 (1964).
- Greer, A. L. *Intermetallic Compounds—Principles and Practice* Vol. 1 (eds Westbrook, J. H. & Fleischer, R. L.) 731–754 (Wiley, New York, 1995).
- Gaskell, P. H. A new structural model for transition metal–metalloid glasses. *Nature* **276**, 484–485 (1978).
- Gaskell, P. H. A new structural model for amorphous transition metal silicides, borides, phosphides and carbides. *J. Non-Cryst. Solids* **32**, 207–224 (1979).
- Elliott, S. R. *Physics of Amorphous Materials* 2nd edn, 139–151 (Longman, London, 1990).
- Gaskell, P. H. *Amorphous Metals* (eds Matyja, H. & Zielinski, P. G.) 35–57 (World Scientific Publishing, Singapore, 1985).
- Gaskell, P. H. Medium-range structure in glasses and low-Q structure in neutron and X-ray scattering data. *J. Non-Cryst. Solids* **351**, 1003–1013 (2005).
- Miracle, D. B. A structural model for metallic glasses. *Nature Mater.* **3**, 697–702 (2004).
- Miracle, D. B., Sanders, W. S. & Senkov, O. N. The influence of efficient atomic packing on the constitution of metallic glasses. *Phil. Mag. A* **83**, 2409–2428 (2003).
- McGreevy, R. L. Reverse Monte Carlo modeling. *J. Phys. Cond. Matter* **13**, R877–R913 (2001).
- Luo, W. K. et al. Icosahedral short-range order in amorphous alloys. *Phys. Rev. Lett.* **92**, 145502 (2004).
- Lamparter, P. Reverse Monte-Carlo simulation of amorphous Ni₈₀P₂₀ and Ni₈₁B₁₉. *Phys. Scr.* **T57**, 72–78 (1995).
- Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A. & Joannopoulos, J. D. Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **64**, 1045–1097 (1992).
- Kresse, G. & Hafner, J. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
- Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* (Clarendon, Oxford, 1989).
- Kresse, G. & Furthmüller, J. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
- Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
- Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
- Wang, Y. & Perdew, J. P. Correlation hole of the spin-polarized electron gas, with exact small-wave-vector and high-density scaling. *Phys. Rev. B* **44**, 13298–13307 (1991).
- Stillinger, F. H. & Weber, T. A. Packing structures and transitions in liquids and solids. *Science* **225**, 983–989 (1984).
- Ankudinov, A. L., Ravel, B., Rehr, J. J. & Conradson, S. D. Real-space multiple-scattering calculation and interpretation of X-ray-absorption near-edge structure. *Phys. Rev. B* **58**, 7565–7576 (1998).
- Finney, J. L. Random packing and the structure of simple liquids. *Proc. R. Soc. A* **319**, 479–493 (1970).
- Finney, J. L. Modeling structures of amorphous metals and alloys. *Nature* **266**, 309–314 (1977).
- Borodin, V. A. Local atomic arrangements in polytetrahedral materials. *Phil. Mag. A* **79**, 1887–1907 (1999).

26. Borodin, V. A. Local atomic arrangements in polytetrahedral materials. II. Coordination polyhedra with 14 and 15 atoms. *Phil. Mag. A* **81**, 2427–2446 (2001).
27. Pauling, L. The principles determining the structure of complex ionic crystals. *J. Am. Chem. Soc.* **51**, 1010–1026 (1929).
28. O'Keeffe, M. & Navrotsky, A. (eds) *Structure and Bonding in Crystals* (Academic, New York, 1981).
29. Qi, D. W. & Wang, S. Icosahedral order and defects in metallic liquids and glasses. *Phys. Rev. B* **44**, 884–887 (1991).
30. Doye, J. P. K. & Walse, D. J. The effect of the range of the potential on the structure and stability of simple liquids: from clusters to bulk, from sodium to C₆₀. *J. Phys. B* **29**, 4859–4894 (1996).
31. Nelson, D. R. Order, frustration, and defects in liquids and glasses. *Phys. Rev. B* **28**, 5515–5535 (1983).
32. Frank, F. C. & Kasper, J. S. Complex alloy structures regarded as sphere packings. I. Definitions and basic principles. *Acta Crystallogr.* **11**, 184–190 (1958).
33. Watson, R. E. & Bennett, L. H. Crystalline and glassy phases of transition-metal-metalloid systems. *Phys. Rev. B* **43**, 11642–11652 (1991).
34. Egami, T. & Waseda, Y. Atomic size effect on the formability of metallic glasses. *J. Non-Cryst. Solids* **64**, 113–134 (1984).
35. Hafner, J. Theory of the formation of metallic glasses. *Phys. Rev. B* **21**, 406–426 (1980).
36. Clarke, A. S. & Jónsson, H. Structural changes accompanying densification of random hard-sphere packings. *Phys. Rev. E* **47**, 3975–3984 (1993).
37. Jónsson, H. & Anderson, C. Icosahedral ordering in the Lennard-Jones liquid and glass. *Phys. Rev. Lett.* **60**, 2295–2298 (1988).
38. Kelton, K. F. *et al.* First X-ray scattering studies on electrostatically levitated metallic liquids: demonstrated influence of local icosahedral order on the nucleation barrier. *Phys. Rev. Lett.* **90**, 195504 (2003).
39. Manoharan, V. N., Elsesser, M. T. & Pine, D. J. Dense packing and symmetry in small clusters of microspheres. *Science* **301**, 483–487 (2003).
40. Caspar, D. L. D. & Klug, A. Physical principles in the construction of regular viruses. *Quant. Biol.* **27**, 1–24 (1962).
41. Zandi, R., Reguer, D., Bruinsma, R. F., Gelbart, W. M. & Rudnick, J. Origin of icosahedral symmetry in viruses. *Proc. Natl Acad. Sci. USA* **101**, 15556–15560 (2004).
42. Greer, A. L. Metallic glasses. *Science* **267**, 1947–1953 (1995).
43. Lamparter, P. Structure of metallic glasses. *Phys. Scr.* **T57**, 45–63 (1995).
44. Kramer, M. J. & Sordet, D. J. Polymorphism in the short-range order of Zr₇₀Pd₃₀ metallic glasses. *J. Non-Cryst. Solids* **351**, 1586–1593 (2005).
45. Dubois, J. M., Gaskell, P. H. & Le Caer, G. A model for the structure of metallic glasses based on chemical twinning. *Proc. R. Soc. A* **402**, 323–357 (1985).
46. Ohkubo, T. & Hirotsu, Y. Electron diffraction and high-resolution electron microscopy study of an amorphous Pd₈₂Si₁₈ alloy with nanoscale phase separation. *Phys. Rev. B* **67**, 094201 (2003).
47. Kob, W. & Andersen, H. C. Testing mode-coupling theory for a supercooled binary Lennard-Jones mixture: The van Hove correlation function. *Phys. Rev. E* **51**, 4626–4641 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by US DoE-BES, with computational resources provided by NERSC. We also thank D. B. Miracle for discussions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.W.S. (hwsheng@jhu.edu) or E.M. (ema@jhu.edu).

ARTICLES

Policing stabilizes construction of social niches in primates

Jessica C. Flack^{1,2,3}, Michelle Girvan¹, Frans B. M. de Waal^{2,3} & David C. Krakauer¹

All organisms interact with their environment, and in doing so shape it, modifying resource availability. Termed niche construction, this process has been studied primarily at the ecological level with an emphasis on the consequences of construction across generations¹. We focus on the behavioural process of construction within a single generation, identifying the role a robustness mechanism²—conflict management—has in promoting interactions that build social resource networks or social niches. Using ‘knockout’ experiments on a large, captive group of pigtailed macaques (*Macaca nemestrina*), we show that a policing function, performed infrequently by a small subset of individuals³, significantly contributes to maintaining stable resource networks in the face of chronic perturbations that arise through conflict. When policing is absent, social niches destabilize, with group members building smaller, less diverse, and less integrated grooming, play, proximity and contact-sitting networks. Instability is quantified in terms of reduced mean degree, increased clustering, reduced reach, and increased assortativity. Policing not only controls conflict^{3–5}, we find it significantly influences the structure of networks that constitute essential social resources in gregarious primate societies. The structure of such networks plays a critical role in infant survivorship⁶, emergence and spread of cooperative behaviour⁷, social learning and cultural traditions⁸.

We operationalize the social niche in graphical terms as the local connections of a node (that is, an individual) in multiple, overlapping social networks and we define social organization as the union of all social niches (Supplementary Fig. 1S). Whereas the ecological niche is composed of resource vectors⁹ (availability of wood for building dams, prey items, and so on), the social niche is composed of an individual's vector of behavioural connections in the set of overlapping social networks in which it participates (Supplementary Fig. 1S). As with ecological niches, social niches vary in quality and affect one another: if one niche fragments, connectivity, and consequently social resource availability in other niches, is affected. The construction of stable social niches requires that individuals have time and security to engage in social interactions.

In primate societies frequent conflict threatens to destabilize social networks and robustness mechanisms have evolved to stabilize within-group behavioural interactions^{2,10}. We hypothesize that third-party policing—physically impartial intervention into conflicts, made possible by the structure of a status communication network³—stabilizes social niches (Fig. 1) (in addition to directly affecting aggression and conflict levels²), allowing group members to interact with a larger, more diverse set of well-connected partners.

Network comparisons and properties

By perturbing the status-signalling network and disabling policing, we quantify the effects of policing on the structure of four social networks: grooming, play, contact-sitting and proximity. We assess four network properties in three conditions: an observed control (OC) condition in which all individuals were present in the group; an experimental knockout condition (EK) in which all individuals except three high-status nodes (see below) were present; and a topological knockout (TK) condition, consisting of the OC condition with three high-status nodes removed from the data (Fig. 2 and Supplementary Fig. 1S).

The OC–TK–EK comparison combines two traditions that have characterized knockout studies. In cell biology, OC is typically compared to the observed network after experimental knockout (EK)¹¹. In studies of the internet and other kinds of technological networks, OC is typically compared to a virtual network (TK) in

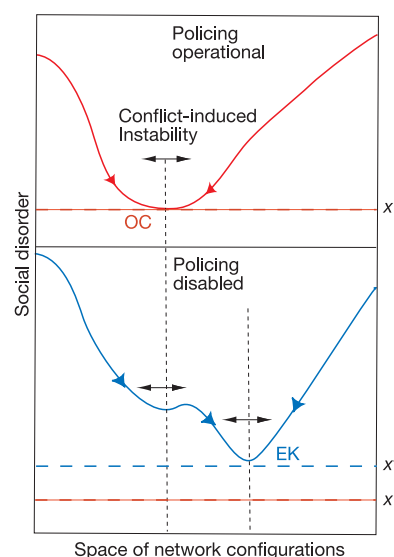


Figure 1 | Schematic showing putative basins of attraction for pigtailed macaque social networks. Policing stabilizes social networks in the OC condition by preventing chronic low-level conflict leading to network fragmentation (level of order indicated by x). When policing is disabled, conflict is no longer moderated, and the social system shifts to a new network configuration, the EK condition, associated with an elevated level of social disorder (indicated by x').

¹Santa Fe Institute, Santa Fe, New Mexico 87501, USA. ²Living Links Center, Yerkes National Primate Research Center, ³Neuroscience and Animal Behavior Program, Emory University, Atlanta, Georgia 30322, USA.

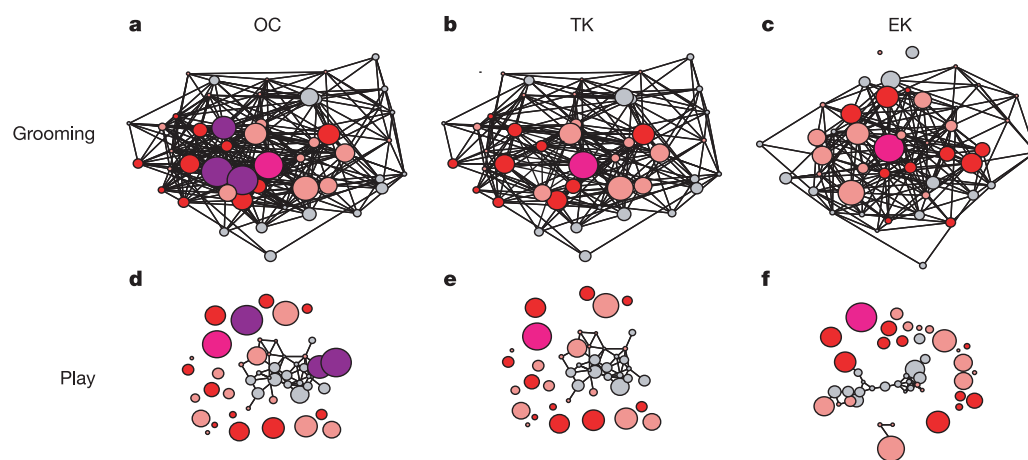


Figure 2 | Empirically derived grooming and play networks in three conditions. OC, all data included; TK, policer data were removed from OC data; EK, policers physically removed from the group. **a**, OC grooming; **b**, TK grooming; **c**, EK grooming; **d**, OC play; **e**, TK play, and **f**, EK play. Node size corresponds to frequency of signals received in the status communication network. Purple nodes: policers. Bright pink node: alpha female. Red nodes: matriarchs. Light pink nodes: other adults. Grey nodes: subadults (socially mature but not fully-grown).

which nodes have been topologically (virtually) removed¹². These comparisons are implicitly considered equivalent. The OC–TK comparison neglects consideration of reconfiguration in the experimental condition. OC–EK does not control for the structural contribution of knockout nodes to network topology. In contrast, TK–EK highlights the extent to which knockout-induced changes in interactions result from simple topological changes. Significant differences between TK and EK reflect behavioural reconfiguration and highlight the systemic role of knockout nodes in network structure.

Knockout node characteristics

The knockout nodes (Fig. 2 and Supplementary Fig. 1S) correspond to three fully grown adult males receiving a disproportionate number of subordination signals, called silent-bared teeth displays (SBT), from 45 mature (84 total) group members in peaceful contexts (for the description of the species studied and the removal procedure see Methods and Supplementary Methods). These signals are unidirectional, that is, they are always emitted by the same individual in the dyad^{13,14}. Peacefully emitted SBTs are considered an acknowledgement of power³. Because they are unidirectional and emitted relatively infrequently, power structure changes slowly.

The observed distribution of SBTs received was not significantly different from log-normal (Kolmogorov–Smirnov test, $P = 0.635$). Individuals receiving many signals were in the distribution's tail. They performed the vast majority of policing interventions terminating conflicts, yet did so rarely compared to the total number of conflicts³. Removal of nodes with heavily weighted input edges (the three high-status nodes) in the signalling network eliminates powerful policers. We have shown elsewhere that policer removal results in organizational destabilization over short timescales as mean levels of aggression increase and mean levels of affiliation, including reconciliation, decrease². Removal, or knockout of highly connected nodes, in systems as diverse as the internet and yeast regulatory networks, has been shown to affect network functionality significantly¹⁵.

Here we investigate how knockout affects four network properties:

mean degree, reach, assortative mixing and clustering. Each of these statistics provides a different insight into the role of policing in social network construction, and differentially emphasizes local versus nonlocal interactions.

Degree results

Degree distributions of grooming, contact-sitting, and proximity networks were normally distributed in OC, TK and EK conditions. Play was right-skewed in each condition. We used repeated measures¹⁶ to analyse how mean degree (node degree is defined as the number of nodes to which it is connected) changed in the networks across the three comparisons (Table 1). TK–EK results indicate that individuals had significantly more play and grooming partners in the presence of policing. Having more partners increases partner choice and redundancy, which is important if partners vary in availability and quality, as has been suggested by work on biological markets¹⁷.

Reach results

Node reach is a measure of its indirect connectedness to other nodes in the graph¹⁸. Here we define node reach to be the number of nodes two or fewer steps away. Reach is important to primates because it affects behavioural contagion. If A grooms B, this can induce B to groom C¹⁹. In this way, positive behaviour propagates over the network. Reach in social networks will also affect contagion of aggression—as a consequence of affiliative relations²⁰, a fight erupting between X and A can cause a chain reaction in which B supports A and C supports B. We consider only two or fewer steps because A's choice of B is more probably dependent on B's actual partners (B's degree) than on more socially distant individuals (B's reach). Additionally, long contagion cascades are more likely to be interrupted by extrinsic factors.

We quantify reach in three of four networks (see Fig. 3a legend). We identify the extent of reach beyond that which would be predicted by degree alone (high degree generally implies high reach). For each network in each condition we generated an ensemble of graphs conserving the original degree distribution (Supplementary Notes

Table 1 | Degree and clustering coefficient repeated measures results

Network	OC-TK		OC-EK		TK-EK	
	Degree	Clustering	Degree	Clustering	Degree	Clustering
Grooming, $n = 45$	SIG D, $P < 0.001$	SIG D, $P < 0.001$	SIG D, $P < 0.001$	SIG D, $P = 0.008$	SIG D, $P = 0.009$	D, $P = 0.36$
Play, $n = 29$	D, $P = 0.08$	EQUAL, $P = 0.94$	SIG D, $P = 0.016$	D, $P = 0.87$	SIG D, $P = 0.027$	D, $P = 0.86$
Contact-sitting, $n = 45$	SIG D, $P < 0.001$	I, $P = 0.028$	D, $P = 0.16$	D, $P = 0.99$	D, $P = 1.0$	D, $P = 0.70$
Proximity, $n = 45$	SIG D, $P < 0.001$	SIG D, $P < 0.001$	D, $P = 0.051$	D, $P = 1.0$	I, $P = 0.051$	SIG I, $P = 0.03$

OC-TK comparison illustrates how networks change structurally following topological removal of policer data, but does not allow for reconfiguration. OC-EK comparison assesses network change following actual policer removal, thereby allowing for reconfiguration, but does not control for structural contribution of policers. TK-EK comparison allows for reconfiguration and controls for policers' direct contribution to network structure. SIG D, significant decrease in mean; SIG I, significant increase in mean; D, mean decreases; I, mean increases; EQUAL, means are equal to two decimal places. Play was tested separately from other variables because it was based on a different sample size. We correct P values for multiple comparisons using the Sidak correction¹⁶. Full results are reported in the Supplementary Data.

and Supplementary Table 1S). We constructed a curve of reach versus degree to reflect the ensemble average. We summed the difference between the reach of each node in the empirical network and the average reach value for a node of that degree in the randomizations and then divided by the total number of nodes. This gave an estimate of average reach deviation from random for each network (Fig. 3a).

For all graphs, reach was less than random expectation given the network's degree profile. Opportunity for contagion was therefore less than expected for a random network. The networks were characterized by relatively low reach in OC, in which direct connections of all nodes, including policers, were considered. In TK (without policers' direct connections), we found policing maintained relatively high potential for contagion (compared to EK) through connections among nonpolicing individuals. The direct connections of policers therefore imparted substantial structure to the graphs. Relative reach in consequence appeared low in OC, but in actuality was high among nonpolicers. This means that policing, through its effects on network structure, maintains potential for both costly aggressive and beneficial affiliative contagion. However, policing

interventions directly control aggressive contagion. Consequently, when policing is operational, it makes sense for nonpolicers to build networks supporting contagion. This interpretation is further supported by low reach in EK combined with increased mean aggression and decreased mean affiliation², suggesting that when policing is absent, individuals do in fact reduce reach to prevent uncontrollable aggressive contagion.

Assortativity results

Networks display assortative mixing by degree when nodes of a given degree attach preferentially to nodes of similar degree. We calculate assortativity after Newman²¹. Assortative mixing gives insight into higher-order structural effects that cannot be gleaned from degree or reach. When individuals interact with others of like degree, they interact with others with access to similar social resources. This is thought to be important for emergence of cooperation²². However, assortativity also results in less interaction-partner diversity and therefore is not an ideal solution for promoting cooperation. As with reach, we controlled for influence of degree profile. As with reach, OC had highest assortativity for all graphs: policers' direct connections contributed structure to the graphs (Fig. 3b). Grooming, play and contact-sitting networks were relatively less assortative in TK than EK. We have shown elsewhere² that cooperation is greater in TK than EK.

Together, these results suggest that policing promotes cooperation among individuals with unequal access to social resources, and also facilitates interaction-partner diversity. When policing was absent (EK), individuals increased assortativity, modifying partner choice to interact with individuals of similar degree. This allowed recovery of some cooperation potential, but at the cost of interaction-partner diversity.

Clustering results

The local clustering coefficient²³ of a node i is the density of its open neighbourhood, where $C_i = \frac{\text{number of triangles connected to } i}{\text{number of triples centred on } i}$. The clustering coefficient for the whole network is given by: $C = \frac{1}{n} \sum_i C_i$. This measure of network transitivity expresses the likelihood that two neighbours of a node will themselves be neighbours. We used repeated measures to investigate the effect of the condition (OC, TK or EK) on clustering in each network (Table 1). Only proximity was significantly affected by policing knockout. Individuals showed more clustering when policing was absent. With policing, individuals exhibited less conservative interactions—they preferentially interacted with more socially distant individuals than with friends of their friends. Policing promotes a more open, integrated society rather than one made up mainly of cliques.

Effect of policing on social niche construction

Conflict threatens to destabilize society. Its immediate consequences include injuries and damaged relationships¹⁰. It has been demonstrated that policing can directly prevent this^{3,4}. We find that policing also has far-reaching indirect consequences, significantly altering construction of social resource networks that make group living advantageous. We demonstrated this by analysing changes to four network properties. We observe that when policing is operational, group members build larger social networks characterized by greater partner diversity and increased potential for socially positive contagion and cooperation. Without policing, high conflict frequency and severity leads to more conservative social interactions and a less integrated society. Mechanisms for buffering frequent conflicts are therefore essential for construction of stable social niches upon which individuals depend for behavioural resources.

METHODS

Pigtailed macaques are indigenous to south East Asia and live in multimale, multifemale societies characterized by female matrilineal and male group transfer upon onset of puberty²⁴. Pigtailed macaques breed all year. Females develop

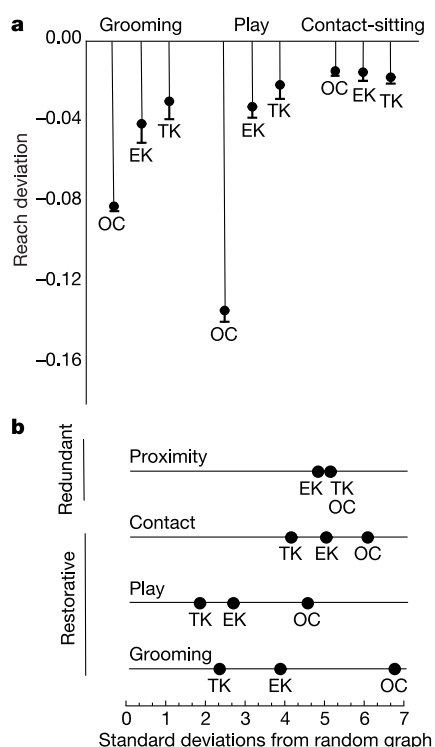


Figure 3 | Reach and assortativity results. **a**, Social network reach. Line segment length represents mean difference (over all nodes) in reach values between the empirical network and the expected reach of a node of identical degree in the randomized networks (500 randomizations). The line segment bars (heavy black lines) indicate the standard deviation of the reach deviation from the mean across all individual randomizations. (A reach deviation from the mean can be calculated for each individual randomization in the same way for the experimental data.) Relative reach in consequence appeared low in OC, but in actuality was relatively high among nonpolicers (TK) compared to EK. Proximity networks were not studied because reach saturated to one for all nodes. All reach deviations are negative because empirical networks display less reach than their randomized counterparts. **b**, Social network assortativity. An ensemble of randomized graphs is constructed for each network (play OC, play TK, play EK, grooming OC, and so on). The measured assortativity of the original graph is compared with the average observed in the ensemble, and measured in units of the standard deviation of the ensemble's assortativity values. The assortativity measures of the graphs are higher than the ensemble average in all cases. Play and grooming are most strongly perturbed by policer knockout. Contact-sitting, play and grooming restore some assortativity through a modification of behaviour by remaining group-members, suggesting a tendency towards behavioural homeostasis.

swellings when in oestrus. The social systems of captive macaque species are relatively well studied²⁵.

Data were collected from a captive, breeding group of pigtailed macaques at Yerkes National Primate Research Center near Lawrenceville, Georgia. The group comprised 84 individuals, including four adult males, 25 adult females, and 19 subadults. All individuals, except eight (four males, four females), were either natal to the group or had been in the group since formation in 1985. The knockout males correspond to three of four males. The group was housed in an indoor–outdoor facility, the outdoor compound of which was 125 feet by 65 feet. During observations, the group was confined to the outdoor area, where all individuals were visible to the observer.

Observation conditions. Observations occurred for up to eight hours daily between 1,100 and 2,000 hours over a twenty-week period from June until October 1998 and were evenly distributed. Provisioning occurred before observations, and once during observations. Data were collected in two conditions: OC (156 h) and EK (78 h). During OC, all individuals were present in the group. During EK, three adult males were simultaneously removed (Supplementary Methods) for the day. OC observations were collected only on days during the twenty-week period when no manipulations to the group occurred. This ‘repeated removal’ design allowed us to control for fluctuations resulting from variation in environmental variables, such as temperature and human activity. Mean temperature was very similar in both conditions (OC, 87.9° F; EK, 89.6° F). On EK days, the group was observed for eight of ten removal hours. Observations began two hours after removal to ensure that stress induced by the benign removal procedure subsided, and did not account for observed changes to networks.

The males were partially removed (limited physical, visual, vocal access to group) on randomly chosen days every two weeks (Supplementary Methods). Removal was temporary (10 h each time) so we assessed whether social networks restructure over short timescales. The brief, partial removal made it possible to isolate effects of policing by providing insurance that observed changes to social networks did not result from an increase in competition over rank vacancies²⁶. As a precaution we evaluated elsewhere whether knockout caused social organization to reconfigure by inducing instability in the dominance hierarchy. We found no support for this hypothesis². We evaluated whether there was redundancy in the system, in that the absence of policers stimulated an increase in policing by other individuals. This was not the case². In Supplementary Data we report results of a control analysis in which a low-ranking female was removed to evaluate whether removal of any individual would negatively affect network robustness. This hypothesis was not supported.

Policers. We evaluated whether changes to social networks result from perturbing affiliation functions performed by policers in addition to their conflict-management functions. The policers were among the strongly connected nodes in three of four social networks. However, their positions in the degree distributions, which (except for play network) were normal and not marked by high variance, suggest that they were not unique. Correspondingly, they did not receive a disproportionate frequency of grooming, as they did status signals (Supplementary Data). Nor did they play ‘broker’²⁷ roles in the social networks—they were not important links connecting relatively isolated clusters of individuals. There was no Pearson correlation between node betweenness²⁸ and node SBT in degree for grooming ($P = 0.88$), contact-sitting ($P = 0.29$) or proximity ($P = 0.31$). These facts increase the likelihood that only conflict-management functionality was disrupted by removal. We note that the objective of the above analyses is to assess the likelihood that policers performed special affiliation functions in addition to conflict-management functions, not to evaluate their direct, structural contributions to the networks, which are controlled for through the three comparisons. The four networks were treated as binary undirected (symmetrical) graphs.

Data sampling methods. Instantaneous scan sampling²⁹ occurred every 15 min for ‘state’ behaviours (grooming, contact-sitting, social proximity and play). All data were collected using a digital stopwatch and voice recorder. 494 OC scans and 235 EK scans were collected. We randomly sampled scans in the OC condition until we had the same number of scans in both conditions (235). F.B.M.d.W. trained the observer (J.C.F.). α (level of significance) was set to 0.05 for repeated measures analyses.

Operational definitions of behaviours. The ‘state’ behaviours are defined as follows: grooming means passing hands or teeth through hair of another individual or plucking the hair with hands or teeth for at least five seconds. Contact-sitting means two or more individuals sitting in contact for at least 5 s. Proximity means that two or more individuals sit within arm’s reach for at least 5 s. Social play is wrestling, hitting, pinching, slapping and chasing, characterized by relaxed muscle movements and involving two or more participants.

Received 3 August; accepted 11 October 2005.

1. Odling-Smee, F. J., Laland, K. N. & Feldman, M. W. *Niche Construction: The Neglected Process in Evolution* (Princeton Univ. Press, Princeton, New Jersey, 2003).
2. Flack, J. C., Krakauer, D. C. & de Waal, F. B. M. Robustness mechanisms in primate societies: A perturbation study. *Proc. R. Soc. Lond. B* **272**, 1091–1099 (2005).
3. Flack, J. C., de Waal, F. B. M. & Krakauer, D. C. Social structure, robustness, and policing cost in a cognitively sophisticated species. *Am. Nat.* **165**, E126–E139 (2005).
4. Frank, S. Repression of competition and the evolution of cooperation. *Evolution* **57**, 693–705 (2003).
5. Boehm, C. in *Primate Behavior and Sociobiology* (eds Chiarelli, A. B. & Corruccini, R. S.) 161–182 (Springer, New York, 1981).
6. Silk, J., Alberts, S. C. & Altmann, J. Social bonds of female baboons enhance infant survival. *Science* **302**, 1231–1234 (2003).
7. Watts, D. J. *Small Worlds: The Dynamics of Networks Between Order and Randomness* 199–222 (Princeton Univ. Press, Princeton, New Jersey, 1999).
8. Whiten, A., Horner, V. & de Waal, F. B. M. Conformity to cultural norms of tool use in chimpanzees. *Nature* **437**, 737–740 (2005).
9. Hutchinson, G. E. Concluding remarks. *Coldspring Harbor Symp. Quant. Biol.* **22**, 415–427 (1957).
10. de Waal, F. B. M. Primates—A natural heritage of conflict resolution. *Science* **289**, 586–590 (2000).
11. Krakauer, D. C. Genetic redundancy. In *Encyclopedia of the Human Genome* Vol. 2, 892–897 (Nature Publishing Group/Wiley, London, 2003).
12. Albert, R., Jeong, H. & Barabasi, A. L. Error and attack tolerance in complex networks. *Nature* **406**, 378 (2000).
13. de Waal, F. B. M. & Luttrell, L. The formal hierarchy of rhesus monkeys: An investigation of the bared-teeth display. *Am. J. Primatol.* **9**, 73–85 (1985).
14. Flack, J. C. & de Waal, F. B. M. in *Macaque Societies: A Model for the Study of Social Organization* (eds Thierry, B., Singh, M. & Kaumanns, W.) 157–181 (Cambridge Univ. Press, Cambridge, UK, 2004).
15. Jeong, H., Mason, H. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
16. Keppel, G. *Design and Analysis: A Researcher's Handbook* 329–415 (Prentice Hall, New Jersey, 1991).
17. Noe, R. & Hammerstein, P. Biological markets. *Trends Ecol. Evol.* **10**, 336–339 (1995).
18. Wasserman, S. & Faust, K. *Social Network Analysis* 107–108 (Cambridge Univ. Press, Cambridge, 1994).
19. de Waal, F. B. M. Attitudinal reciprocity in food sharing among brown capuchin monkeys. *Anim. Behav.* **60**, 253–261 (2000).
20. de Waal, F. B. M. & Luttrell, L. Mechanisms of social reciprocity in three primate species: symmetrical relationships characteristics or cognition? *Ethol. Sociobiol.* **9**, 101–118 (1988).
21. Newman, M. E. J. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003).
22. Trivers, R. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
23. Watts, D. J. & Strogatz, S. H. Collective dynamics of small-world networks. *Nature* **393**, 1302–1305 (1998).
24. Caldecott, J. O. *An Ecological and Behavioral Study of the Pigtailed Macaque* (Contributions to Primatology, S. Karger, Basel, 1986).
25. Thierry, B. in *Natural Conflict Resolution* (eds Aureli, F. & de Waal, F. B. M.) 106–128 (Univ. California, Berkeley, 2000).
26. de Waal, F. B. M. The organization of agonistic social relationships within two captive groups of Java-monkeys (*Macaca fascicularis*). *Z. Tierpsychol.* **44**, 225–282 (1977).
27. Lusseau, D. & Newman, M. Identifying the role that individual animals play in their social network. *Proc. R. Soc. Lond. B* **271**, S477–S481 (2004).
28. Freeman, L. C. A set of measures of centrality based upon betweenness. *Sociometry* **40**, 35–41 (1977).
29. Altmann, S. Observational study of behaviour: sampling methods. *Behaviour* **49**, 227–267 (1974).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Jen, E. Goldberg, J. Miller, E. Smith, D. Erwin, L. Marino, H. Gouzoules, C. Boehm, N. Ay, Y. Sato, M. Morris and J. Padgett for discussions, M. Seres for assistance with the data collection, and the animal care staff at YNPRC. YNPRC is fully accredited by the American Association for Accreditation of Laboratory Animal Care. This work was supported by the Packard Foundation (D.C.K.), the McDonnell Foundation (D.C.K., J.C.F., M.G.), the Wenner-Gren Foundation (J.C.F.), the Proteus Foundation (D.C.K.), the NSF (F.B.M.d.W.) and the NIH (F.B.M.d.W., J.C.F., D.C.K.).

Author Information Reprints and permissions information is available at ngp.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.C.F. (jflack@santafe.edu).

ARTICLES

Epigenetic silencers and Notch collaborate to promote malignant tumours by *Rb* silencing

Dolors Ferres-Marco^{1*}, Irene Gutierrez-Garcia^{1*}, Diana M. Vallejo^{1*}, Jorge Bolivar², Francisco J. Gutierrez-Aviño¹ & Maria Dominguez¹

Cancer is both a genetic and an epigenetic disease. Inactivation of tumour-suppressor genes by epigenetic changes is frequently observed in human cancers, particularly as a result of the modifications of histones and DNA methylation. It is therefore important to understand how these damaging changes might come about. By studying tumorigenesis in the *Drosophila* eye, here we identify two Polycomb group epigenetic silencers, Pipsqueak and Lola, that participate in this process. When coupled with overexpression of Delta, deregulation of the expression of Pipsqueak and Lola induces the formation of metastatic tumours. This phenotype depends on the histone-modifying enzymes Rpd3 (a histone deacetylase), Su(var)3-9 and E(z), as well as on the chromodomain protein Polycomb. Expression of the gene *Retinoblastoma-family protein (Rbf)* is downregulated in these tumours and, indeed, this downregulation is associated with DNA hypermethylation. Together, these results establish a mechanism that links the Notch-Delta pathway, epigenetic silencing pathways and cell-cycle control in the process of tumorigenesis.

Correct organ formation depends on the balanced activation of conserved developmental signalling pathways (such as the Wnt, Hedgehog and Notch pathways). If insufficient signals are received, organ growth may be deficient. By contrast, excess signalling leads to an overproduction of progenitor cells and a propensity to develop tumours^{1–3}. When such hyperproliferation is associated with the capacity of cells to invade surrounding tissue and metastasis to distant organs, cancer develops⁴. Indeed, activation of the Wnt, Hedgehog and Notch pathways is a common clinical occurrence in cancers^{5,6}. Curiously, activation of any of these pathways in animal models seems to be insufficient for cancer to develop, indicating that synergism with other genes is required for these pathways to produce cancer.

Cellular memory or the epigenetic inheritance of transcription patterns has also been implicated in the control of cell proliferation during development, as well as in stem-cell renewal and cancer^{7,8}. Proteins of the Polycomb group (PcG) are part of the memory machinery and maintain transcriptional repression patterns⁹. The upregulation of several PcG proteins has been associated with invasive cancers⁸. Thus, increased amounts of EZH2, the human homologue of the *Drosophila* histone methyltransferase E(z)¹⁰, is associated with poorer prognoses of breast and prostate cancers⁸.

Another histone methyltransferase implicated both in gene silencing and in cancer is SUV39H1, a homologue of *Drosophila* Su(var)3-9 (ref. 11). SUV39H1 and Su(var)3-9 methylate histone H3 on lysine 9 (H3K9me), and this epigenetic tag is characteristic of heterochromatin and DNA sequences that are constitutively methylated in normal cells¹². DNA methylation is another mechanism involved in cellular memory that actively contributes to cancer^{7,13}. Indeed, numerous tumour-suppressor genes, including the retinoblastoma gene *RB*, are silenced in cancer cells by DNA

hypermethylation¹³. Inactivation of the *RB* tumour-suppressor pathway is considered an important step towards malignancy¹⁴; thus, it is important to understand how these damaging epigenetic changes are initiated in cells that become precursors of cancer. Moreover, it is equally important to determine the connection between these processes and the developmental pathways controlling proliferation.

Forward genetic screening in suitable animals is a powerful tool with which to identify tumour-inducing genes and to reveal changes that precede neoplastic events *in vivo*. The developing eye of *Drosophila melanogaster* is a good model for such studies¹⁵ because it is a simple and genetically well-defined organ. The growth of the eye depends on Notch activation in the dorsal-ventral organizer by its ligands Delta (human counterparts, DLL-1, -3, -4) and Serrate (human counterparts, JAGGED-1, -2)^{3,16}. Here we used the 'large eye' phenotype, produced by overexpression of *Delta*, as a tool to screen for mutations that interact with the Notch pathway and convert tissue overgrowths into tumours (Fig. 1a). We isolate one mutation, *eyeful*, that combined with Delta induces metastatic tumours. *eyeful* forces the transcription of two hitherto unsuspected growth and epigenetic genes, *lola* and *pipsqueak* (*psq*). The identification of *eyeful* has been a starting point from which to unravel crosstalk between the Notch and epigenetic pathways in growth control and tumorigenesis. The fact that many epigenetic factors are involved in cancer suggests that these processes may be more generally involved in tumorigenesis than at first it might seem.

Genetic screen and isolation of *eyeful*

To identify genes that interact with the Notch pathway and that influence growth and tumorigenesis, we used the Gene Search (GS) system¹⁷ to screen for genes that provoked tumours when co-expressed with Delta in the proliferating *Drosophila* eye (Fig. 1a).

¹Instituto de Neurociencias CSIC-UMH, Campus de San Juan, Apartado 18, 03550 Sant Joan, Alicante, Spain. ²Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad de Cadiz, Polígono Río San Pedro s/n, 11510 Puerto Real, Spain.

*These authors contributed equally to this work.

The *ey-Gal4* line was used for both eye-specific and ubiquitous induction, resulting in the transactivation of UAS-linked genes throughout the proliferating eye discs (Fig. 1b). It was through such a screen that we isolated the GS88A8 line. Generalized overexpression of *Delta* by *ey-Gal4* (hereafter termed *ey-Gal4 > Dl*) produces mild eye overgrowth¹⁸ (Fig. 1c, d). In most of the flies in which the GS88A8 line was coexpressed with *Delta*, tumours developed in the eyes (Fig. 1e, f). Moreover, in ~30% of the mutant flies, secondary eye growths were observed throughout the body (Fig. 1e–i), and in some flies the whole body filled up with eye tissue. These secondary eye growths had ragged borders, indicating invasion of the mutant tissue into the surrounding normal tissue (Fig. 1g–i). As a result, we named the GS88A8 line ‘*eyeful*’.

We next undertook a developmental analysis of the tumours (Fig. 2). To facilitate analyses, we generated a triple mutant strain carrying the *eyeful*, *UAS-Dl* and *ey-Gal4* transgenes all on the same chromosome (*ey-Gal4 > eyeful > Dl*; see Supplementary Information). In this strain, mutant eye discs showed massive uncontrolled overgrowth (some discs were more than five times their normal size; Fig. 2b–d). In most discs, the epithelial cells had lost their apical–basal polarity (Fig. 2c), and some had a disrupted basement membrane (Fig. 2d) and grew without differentiating.

We extended these results to the wing disc. First, we used *dpp-Gal4* to direct coexpression of *eyeful* and *Delta* along the anterior–posterior boundary of the wing (perpendicular to the endogenous *Delta* domain along the dorsal–ventral boundary; Fig. 2e–h). In a normal wing disc, the *dpp-Gal4* driver typically establishes a stripe of green fluorescent protein (GFP) expression with a sharp border at the boundary (Fig. 2e). Whereas wild-type (or single *eyeful*) cells expressing GFP conformed with this pattern (data not shown), some of the *eyeful* and *Delta* cells were found outside this stripe (Fig. 2h), indicating that the mutant cells can disseminate and invade adjacent regions of the disc. Second, the *MS1096-Gal4* line was used to direct expression in the dorsal wing disc compartment. Under these conditions, the wing tissue grew massively and aggressively, and

the mutant tissue failed to differentiate (Fig. 2i). Together with the results in Fig. 1e–i, these observations suggest that, when coupled with *Delta* overexpression, an excess of the gene products flanking the *eyeful* insertion site induces the formation of tumours capable of metastasising.

eyeful* enforces the transcription of *lola* and *psq

We isolated and sequenced the genomic DNA flanking the *eyeful* P-element (Fig. 3a and Supplementary Table S1 and Fig. S1a). *eyeful* is inserted in an intron of the gene *longitudinals lacking* (*lola*), which is known to be a chief regulator of axon guidance^{19–21}. *lola* encodes 25 messenger RNAs that are produced by alternative splicing and that generate 19 different transcription factors. All of the different isoforms share four exons that encode a common amino terminus, which contains a BTB or POZ domain. In addition, all but one of these transcription factors are spliced to unique exons encoding one or a pair of zinc-finger motifs^{19,22}.

The GS P-elements allow Gal4-dependent inducible expression of sequences flanking the insertion site in both directions¹⁷ (Fig. 1a). The nearest gene in the opposite direction to transcription of *lola* is the *psq* gene (Fig. 3a). This gene encodes nine variants produced by alternative splicing and alternative promoter use, generating four different proteins. Three of the *psq* isoforms contain a BTB or POZ domain in the N terminus, and a histidine- and glutamine-rich region downstream of this domain. Two of the BTB-containing isoforms and the isoform that lacks this domain contain four tandem copies of an evolutionarily conserved DNA-binding motif, the Psq helix–turn–helix (HTH) motif^{23–25}.

psq was initially identified for its ‘grandchildless’ and posterior group defects^{23,26} and was subsequently shown to have a role in retinal cell fate determination²⁴. Psq is essential for sequence-specific targeting of a PcG complex that contains histone deacetylase (HDAC) activity⁹. Psq binds to the GAGA sequence^{27,28}, which is present in many Hox genes and in hundreds of other chromosomal sites⁹.

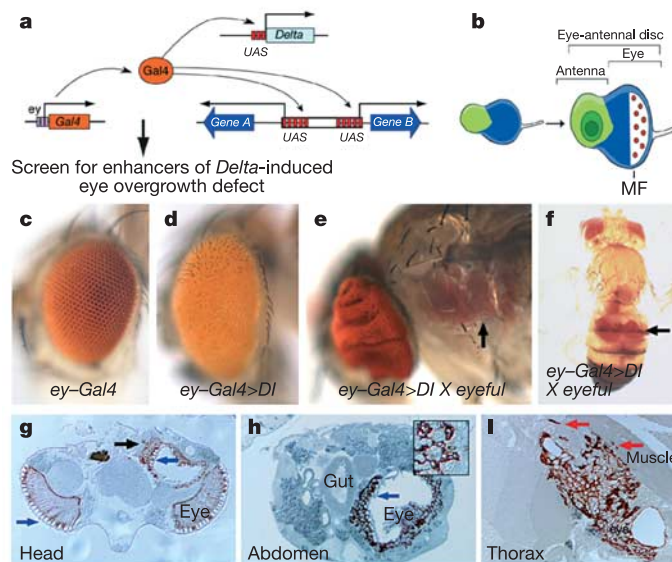


Figure 1 | Isolation of tumour-initiating genes in *Drosophila melanogaster*. **a**, Design of the GS screen for enhancers of the *Delta*-induced eye overgrowth defect. Flies carrying the *ey-Gal4* and *UAS-Delta* (*ey-Gal4 > Dl*) constructs were crossed to randomly inserted GS lines (D.F.-M. and M.D., unpublished data). In these flies, Gal4 binds to the UAS sites and causes coexpression of *Delta* and the gene or genes adjacent to the GS P-element insertion in the developing eye disc. **b**, Schematic of *ey-Gal4* expression (blue) in the cycling cells anterior (to the left) to the morphogenetic furrow (MF) in the larval eye-antennal disc. Postmitotic

retinal differentiating cells (red circles) are indicated. **c**, Control *ey-Gal4* adult eye. **d**, *ey-Gal4 > Dl* adult ‘large eye’. **e**, **f**, Adults from the cross of *eyeful* to *ey-Gal4 > Dl* show massive eyes and secondary eye growths (arrows). **g–i**, Sections through a head (**g**), abdomen (**h**) and thorax (**i**) carrying secondary eye growths. Note the ragged borders of these eye growths (black arrow). Lenses (blue arrows) and dispersed groups of retinal cells (red arrows) are indicated. Inset shows a detail of ommatidial arrangement in these eye growths.

Both polymerase chain reaction with reverse transcription (RT-PCR) and *in situ* hybridization experiments (Supplementary Fig. S1b–g) confirmed that transcription of *lola* and *psq* was influenced by *eyeful* in response to Gal4 activation.

Both genes seem to contribute to the tumour phenotype

To determine whether *lola* and/or *psq* was responsible for the tumour phenotype, we tested 11 enhancer promoter (EP) P-elements inserted into the *lola* and *psq* region (Supplementary Fig. S1a). In contrast to the GS lines, the EP lines allows Gal4-dependent inducible expression of sequences flanking only one end of the P-element²⁹. We found that none of the EP lines induced tumours (Supplementary Table S1); thus, we reasoned that the deregulation of both genes might be required to produce the tumours.

The complexity of *lola* and *psq* loci, which together produce 23 proteins, hampers identification of the transcripts responsible for the *eyeful* phenotype by gain-of-expression mutants (that is, by expressing individual or combinations of isoforms). Therefore, we resolved this issue by isolating point mutations that reverted the phenotype caused by deregulated expression of *lola* and *psq*. In this analysis, the

chemical mutagen ethyl-methane sulphonate (EMS) was used to induce preferentially single nucleotide changes.

The parental *eyeful* GS line was viable in *trans* with deficiencies that removed both *lola* and *psq* (data not shown). In contrast, a set of 14 EMS-induced mutations on the *eyeful* chromosome failed to complement these deficiencies and were found to be alleles of *psq* or *lola* (Fig. 3b–f and Supplementary Figs S2 and S3). We sequenced the EMS-induced mutations that best recovered a normal eye size. Each individual mutation had a single base change or a small deletion that considerably altered the predicted Psq or Lola proteins.

All *psq*[−] mutations induced on the *eyeful* chromosome prevented *eyeful* from producing eye tumours and metastases (Fig. 3b, c, e). Three alleles affected the BTB domain (*psq*^{rev2}, *psq*^{rev7} and *psq*^{rev9}), and three other alleles contained either a premature stop codon that would produce truncated proteins lacking the Psq HTH repeats (*psq*^{rev4} and *psq*^{rev14}) or a missense mutation that would change a conserved amino acid in the third Psq HTH repeat (*psq*^{rev12}). All *lola*[−] mutations induced on the *eyeful* chromosome, including the presumptive null allele (*lola*^{rev6}; Fig. 3d, f), reduced eye tumour size but still permitted sporadic secondary growth (Supplementary Information).

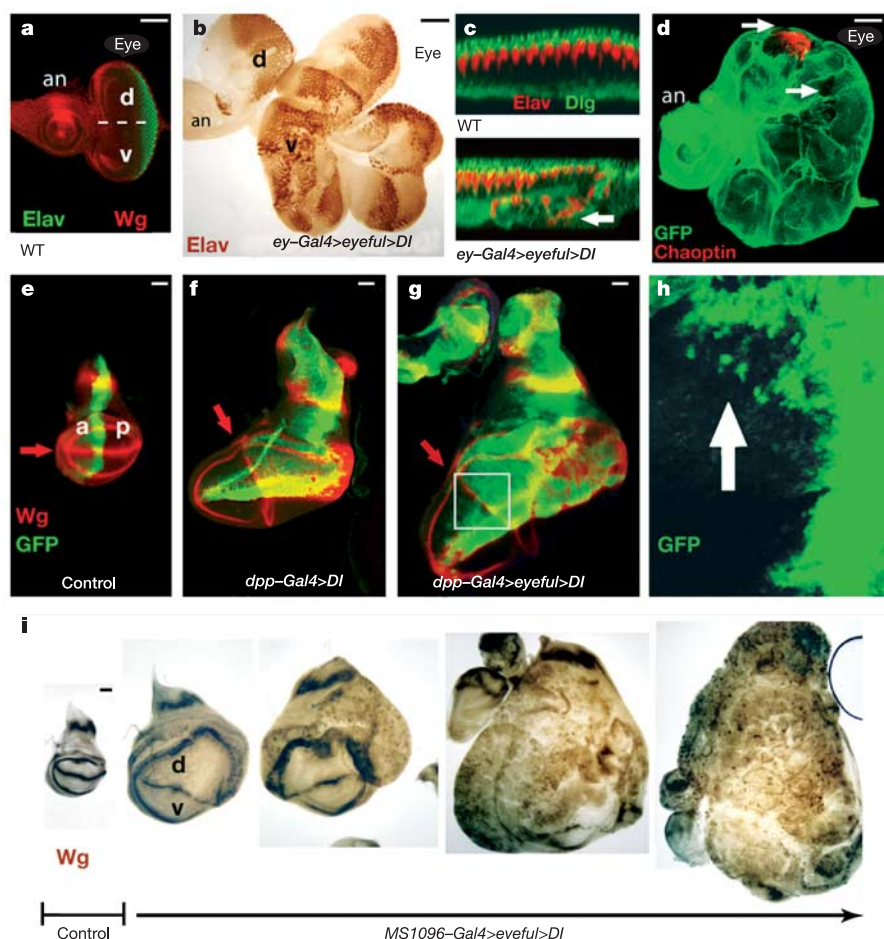


Figure 2 | *eyeful* collaborates with *Delta* to cause malignant tumours.

a–c, Wild-type (**a**) and 7-day-old mutant (**b**) eye discs. Dorsal eye cells are refractory to Delta^{33–35} and overgrowth occurs predominantly in the ventral region. The dorsal–ventral boundary is indicated. Elav marks the nuclei of photoreceptor neurons. **c**, Vertical confocal section of a wild-type (top) and a mutant (bottom) eye disc dissected from a 9-day-old larva. Cell membranes are highlighted by Dlg (green) staining. The disc is a bilayer epithelium comprising the squamous epithelium (or ‘peripodial membrane’) and the main disc epithelium where the nuclei of photoreceptor neurons (red) are apically located. Note the disorganization of the nuclei of neurons (arrow) in the mutants. **d**, Disruption (arrows) of the basement

membrane visualized by GFP-tagged Viking/Collagen IV (green). Retinal differentiation (Chaoptin, red) is apparent only in a dorsal spot. **e**, Control *dpp-Gal4 > GFP* wing disc size. The anterior–posterior boundary is indicated; Wg staining marks the dorsal–ventral boundary (red arrows). **f**, *dpp-Gal4 > DI* wing disc overgrowth. **g**, **h**, *dpp-Gal4 > eyeful > DI* wing disc massive overgrowth (**g**) and invasive migratory cellular behaviour (arrow, **h**). **i**, Evolution of wing disc tumours caused by coexpression of *eyeful* and *Delta* in the MS1096-Gal4 line. Ventral wing cells are refractory to Delta³³; thus, overgrowth in these discs initiates in the dorsal region but progressively extends to occupy the whole disc (all discs are at the same magnification). Scale bars, 50 μ m.

These data show an unequal contribution of Psq and Lola in this process, whereby Psq is the most important factor in the tumorigenic phenotype. The BTB subfamily of transcriptional repressors includes the human oncogenes *BCL6* and *PLZF*. In these oncogenes, the BTB domain is crucial for oncogenesis through the recruitment of PcG and HDAC complexes^{30,31}. We therefore speculated that deregulated Psq and Lola could lead to tumorigenesis by epigenetic processes and that *Drosophila* counterparts of HDACs and PcG proteins might be involved in the progression of these tumours. Indeed, we found genetic evidence that both Lola and Psq function as epigenetic silencers *in vivo* (Supplementary Fig. S4).

Loss of trimethylation of histone H3 on K4 in tumours

We therefore attempted to determine the specific epigenetic mechanisms through which deregulation of Psq and Lola might induce tumorigenesis in conjunction with *Delta* overexpression. Methylation of histone on lysine is a central modification in both epigenetic gene control and in large-scale chromatin structural organization¹². For example, trimethylation of histone H3 on K4 (H3K4me3) is associated with the active transcription of genes and open chromatin structure. By contrast, histone hypoacetylation and H3K9 and H3K27 methylation are characteristic of heterochromatin state and gene silencing^{12,32} (Fig. 4a). To determine whether any changes in these epigenetic markers might coincide with the induction of tumorigenesis, we immunolabelled eye discs with antibodies against specific histone H3 modifications. Because dorsal eye disc cells are refractory to *Delta*^{33–35}, the dorsal region of the discs provided an internal control for these studies.

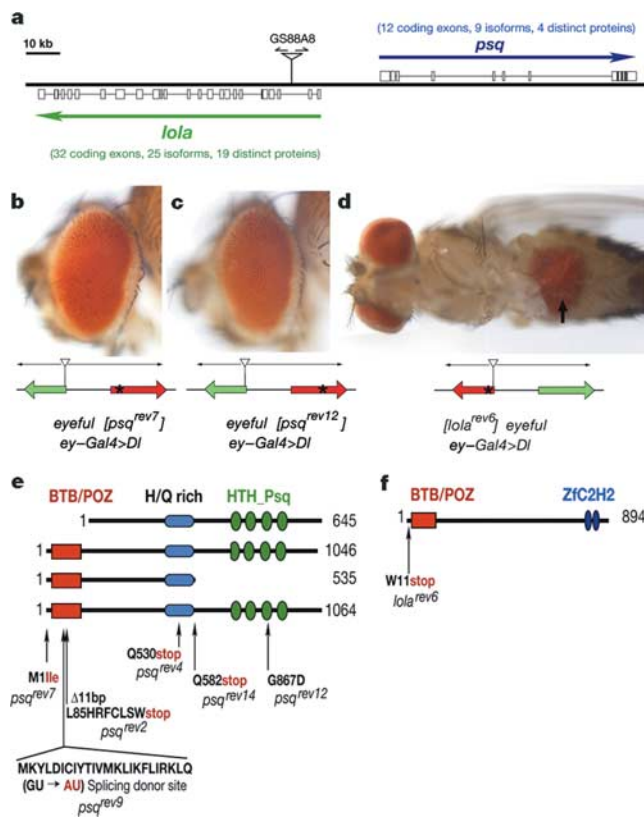


Figure 3 | *eyeful* is a complex mutation affecting two BTB family genes, *lola* and *psq*. **a**, Genomic organization of the *lola* and *psq* genes in the cytological region 47A11-B1 and the *eyeful* GS insertion site. **b–d**, EMS-induced revertants of *eyeful* involve mutations in *psq* (**b**, **c**) or *lola* (**d**). Arrow indicates a large secondary eye growth in the presumably null *lola*^{rev6} revertant. **e**, **f**, Schematic of the four Psq variant proteins (**e**) and a zinc-finger-containing Lola variant protein (**f**) showing the position of the mutations. Numbers refer to the longest, Psq-PB and Psq-PC, variants (**e**).

With the exception of some scattered cells, a prominent loss or strong reduction of H3K4me3 was observed in the ventral region of the mutant discs (Fig. 4b). Notably, although the loss of *Notch* in clones does not affect this epigenetic tag³⁶, overexpression of *Delta* caused a significant reduction in staining for H3K4me3 (Fig. 4b). The H3K4me3 depletion was already apparent in discs showing moderate hyperplasia and thus preceded neoplasm formation. We could not reproducibly resolve changes in other epigenetic tags (such as H3K9me3 and H3K27me2); perhaps more sensitive methods or antibodies might facilitate detection of such changes.

Histone deacetylation and methylation in tumours *in vivo*

H3K4 methylation is thought to be permissive for maintaining and propagating activated chromatin and is thought to neutralize repressor tags by precluding binding of the HDAC complex and impairing SUV39H1-mediated H3K9 methylation¹². Thus, H3K4me3 depletion may contribute to tumour formation by permitting aberrant chromatin silencing (Fig. 4a). We found that a 50% reduction in dosage of the HDAC gene *Rpd3* (Fig. 4c) or of *Su(var)3-9* (Fig. 4d) decreased the tumour phenotype dominantly. In contrast, reducing the activity of the H3K4 histone methyltransferase genes⁹ *Trx* (known as *ALL1* or *MLL* in humans) or *Ash1* (Fig. 4a), which would be expected to deplete the H3K4me3 tag further, did not visibly enhance the tumours (data not shown).

E(z) when complexed with the Extra sex combs (Esc) protein becomes a histone methyltransferase¹⁰. The E(z)–Esc complex and its mammalian counterpart Ezh2–Eed show specificity for H3K27 but may also target H3K9 (refs 10, 12, 37, and Fig. 4a). The complex also contains the HDAC Rpd3, and this association with Rpd3 is conserved in mammals³⁸. H3K27 methylation facilitates binding of the chromodomain protein Pc (HPC in humans)^{37,39}, which

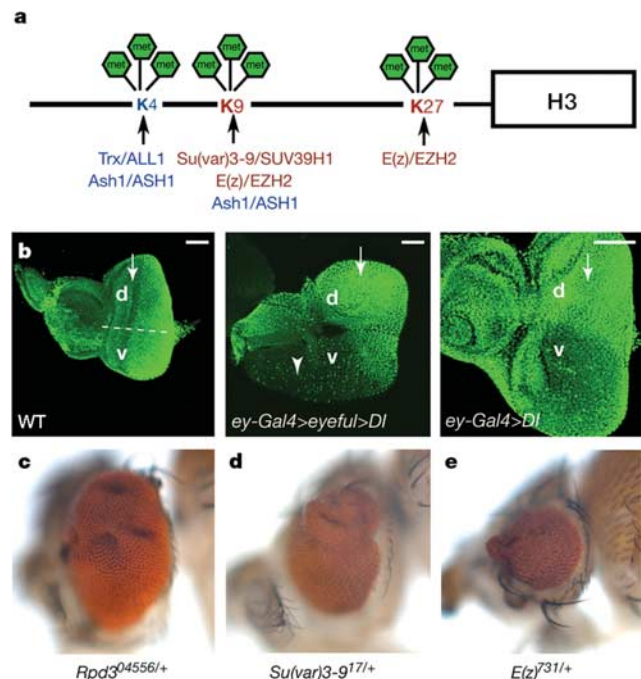


Figure 4 | Aberrant histone modifications associated with the tumours. **a**, Overview of site-selective specificity of histone methyltransferases for distinct lysine (K) positions in the N terminus of histone H3. Histone methyltransferases associated with active chromatin are blue and those associated with silenced chromatin are red. **b**, Wild-type and altered patterns of H3K4me3 in the tumour (*ey-Gal4 > eyeless > Df*) and overgrowth (*ey-Gal4 > Df*) genotypes. The furrow (arrows) and the dorsal and ventral regions of the eye discs are indicated. **c–e**, Representative adult eyes of *ey-Gal4 > eyeless > Df* flies carrying a mutant copy of the HDAC gene *Rpd3* (**c**), *Su(var)3-9* (**d**) and *E(z)* (**e**). Scale bars, 50 μ m.

then creates a repressive chromatin state that is a stable silencer of genes.

Although loss of *E(z)* does not cause proliferation defects within discs^{37,40}, halving the *E(z)* gene dosage dominantly suppressed tumorigenesis (Fig. 4e), indicating that histone methylation by the *E(z)*–*Esc* complex is also a prerequisite for the excessive proliferation

of these tumours. Accordingly, *Esc*[−] or *Pc*[−] mutations also notably reduced the tumours (Supplementary Information).

Together, these findings suggest that the development of these tumours involves, at least in part, changes in the structure of chromatin brought about by covalent modifications of histones. These changes probably switch the target genes from the active H3K4me3 state to a deacetylated H3K9 and H3K27 methylation silent state.

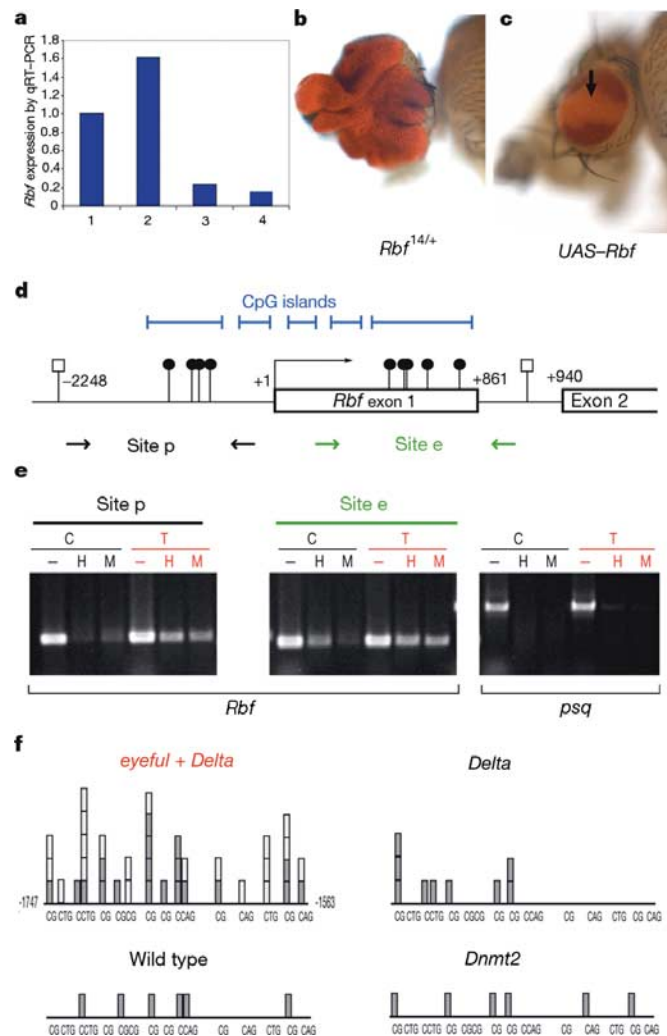


Figure 5 | Silencing of *Rbf* is associated with promoter DNA hypermethylation. **a**, *Rbf* gene expression was analysed by quantitative real-time PCR. The relative levels of *Rbf* normalized to *Rp49* (as a loading control) signals are shown for 5-day-old eye discs from wild type (lane 1), *ey-Gal4 > eyeful* (lane 2), *ey-Gal4 > eyeful > Dl* (lane 3) and *ey-Gal4 > Dl* (lane 4). **b**, A mutated allele of *Rbf* (*Rbf*¹⁴) enhances the tumour defect of *ey-Gal4 > eyeful > Dl* flies. **c**, In contrast, eye-specific overexpression of *UAS-Rbf* rescues the tumours. Arrow indicates transposon silencing (see Supplementary Fig. S4). **d**, CpG-rich islands in the *Rbf* promoter and exon 1. The upstream region of *Rbf* overlaps with a predicted gene named CG13359, which is not conserved in *Drosophila pseudoobscura*. HindIII (squares) and HpaII/MspI (circles) sites are indicated. **e**, Methylation-sensitive restriction digests followed by PCR amplification. PCR yields products only when the region is methylated at each HpaII/MspI site. Shown are representative gels for *Rbf* and *psq* (as a control for a gene expressed in the tumours) of *w*¹¹¹⁸ (C) and *ey-Gal4 > eyeful > Dl* (T). Uncut (−) DNA was used as a loading control and as a control for the PCR reaction. **f**, DNA methylation in the *Rbf* region in mutant and control eye disc cells as determined by bisulphite genomic sequencing (Supplementary Information). The number and position of the methylated cytosines (out of a possible 13) are indicated by filled rectangles (light grey squares denote hemi-methylation).

Silencing of *Rbf* in overgrowth and tumours

From the above data, we considered that the tumours might form as a result of aberrant gene silencing. If so, then the expression of genes involved in cell-cycle control is likely to be altered in the mutant cells. We compared the transcription of 12 tumour-related genes in the mutant and wild-type discs. Transcription of the gene *Rbf*, a fly homologue of the *RB/Rb* family of genes⁴¹, was strongly down-regulated in this assay (and even in *ey-Gal4 > Dl* flies; Fig. 5a). A second *Rb* gene, *Rbf2*, remained unchanged in the different genetic backgrounds, highlighting the specificity of *Rbf* silencing (data not shown).

We found that *Rbf* depletion seems to be intricately associated with tumorigenesis: first, reducing *Rbf* gene dosage by 50% enhanced tumour growth (Fig. 5b); second, re-establishing *Rbf* expression in the eye (using an *UAS-Rbf* transgene) consistently prevented eye tumours and occurrence of secondary growths (in 100 flies examined; Fig. 5c).

Inactivation of *RB1* in retinoblastoma, a form of eye cancer in children, can occur through DNA hypermethylation of the promoter¹³. Unlike in mammals, however, there is little cytosine methylation of the genome in *Drosophila* during developmental stages^{42,43}, and its potential role during tumorigenesis is unknown. DNA methylation seems to depend on one DNA methyltransferase, Dnmt2, that preferentially methylates cytosine at CpT or CpA sites^{44–47}. The fly genome also encodes one methyl-CpG DNA-binding MBD2/3 protein⁴⁸. Because there are no known *Dnmt2* loss-of-function mutations, we could not test the role of this gene in tumorigenesis.

Nevertheless, we tested whether the CpG islands that we observed in the *Rbf* gene were potential targets for repression by DNA methylation by two methods (Fig. 5d). First, we used methylation-sensitive restriction enzymes analysis (Supplementary Information), which showed that the regions around the promoter and transcription start site of the *Rbf* gene were susceptible to methylation (Fig. 5d, e). This approach showed aberrant DNA hypermethylation of *Rbf* in the *eyeful* and *Delta* eye discs (Fig. 5e) and mild hypermethylation in *Delta* discs (data not shown); however, at best only very mild methylation was detected in discs from wild-type flies or from flies with the control *psq* gene (Fig. 5e).

Second, we carried out direct bisulphite sequencing of genomic DNA from mutant discs (Fig. 5f). This approach confirmed the notable increase in methylated DNA in *eyeful* and *Delta* discs when compared with wild-type discs (and a moderate increase in methylated DNA in the *Delta* discs). Hypermethylation of the *Rbf* promoter was not simply the result of *de novo* transcription of *Dnmt2* (*ey-Gal4 > Dnmt2*; Fig. 5f), indicating that activation of the Notch pathway is a crucial step in this *de novo* hypermethylation of *Rbf*.

Discussion

Here we have used *Drosophila* genetics to search for genes that collaborate with the Notch pathway during tumorigenesis *in vivo*. We have identified Psq and Lola as decisive factors to foment tumour growth and invasion when coactivated with the Notch pathway. These proteins are presumptive transcription repressors that contain a BTB domain and sequence-specific DNA-binding motifs and behave as epigenetic silencers *in vivo*.

In addition, we have identified crosstalk between the Notch

pathway and different epigenetic regulators. It is likely that alterations in this crosstalk provoke the aberrant epigenetic repression (and perhaps also derepression) of genes that contributes to cellular transformation. We have identified the *Rbf* gene as one target for this epigenetic regulation and shown that *Rbf* depletion directly contributes to the tumours.

We propose that the sequence of events that leads to these tumours commences with hyperactivation of the Notch pathway, which initiates gene repression. Subsequently, or at the same time as Notch, Psq-Lola could bind to the silenced genes and enforce silencing by recruitment of HDAC or PcG repressors. Given the conservation of the Psq-like HTH domains in Psq and of BTB domains, it seems likely that other transcriptional repressors containing such domains strongly influence the tumour-inducing capacities of HDACs and PcG repressors in human cancers.

Finally, the collaboration between PcG-mediated cellular memory and the Notch pathway may have implications in other processes controlled by Notch, including the second mitotic wave in the *Drosophila* eye^{49,50}, and the organization of eye and wing growth^{3,16}. In these processes, the memory mechanism could ensure that cells kept a record of the Notch signals received at an earlier stage or when the progenitor cells were closer to the Delta source. In this way, they might remain proliferative without having to receive continuous instructions from Notch. Likewise, such a situation could be conceived for tumorigenesis. The oncogenic signals could opportunistically take advantage of the memory mechanism to fix and to maintain their instructions of continuous proliferation in progenitor or stem cells, thereby fostering tumour growth and metastasis.

METHODS

***Drosophila* husbandry and the genetic screen.** The GS88A8 (*eyeful*) and GS71A5 lines and other mutant stocks and transgenes are described in the Supplementary Information.

GS element mapping. Genomic DNA flanking the P-element insertions in the GS88A8 and GS71A5 lines were recovered by inverse PCR (<http://www.fruitfly.org/about/methods>) and sequenced.

***In situ* hybridization and antibody staining.** We carried out *in situ* hybridization and antibody staining as described in Supplementary Information.

EMS reversion mutagenesis and phenotypic analyses. Details of mutagenesis and phenotypic analyses are described in the Supplementary Information.

Real-time PCR. Real-time PCR was done with a BioRad iCycler in the presence of SYBR Green I according to the manufacturer's instructions. The primers of the *Rbf* gene and the control *Rp49* gene were designed with Primer Express software (Applied Biosystems). To assess the expression of *Rbf* and *Rp49* (as a loading control), we extracted total RNA from antennal-eye imaginal discs with Trizol Reagent (Invitrogen). We prepared complementary DNA with an Enhanced Avian RT-PCR kit (Sigma) using oligo (dT)₂₃ primers and 2 µg of total RNA. Expression levels were calculated as a ratio between the signals from *Rbf* and those from *Rp49*. The nature of the PCR products was confirmed by agarose gel electrophoresis and melting curve analysis.

Bisulphite and methylation analysis. Genomic DNA was isolated from third instar larvae or dissected eye discs and the DNA was analysed by methylation-sensitive enzyme digestion and direct bisulphite sequencing as described in the Supplementary Information.

Received 30 June; accepted 20 October 2005.

- Logan, C. Y. & Nusse, R. The Wnt signalling pathway in development and disease. *Annu. Rev. Cell. Dev. Biol.* **20**, 781–810 (2004).
- Hooper, J. E. & Scott, M. P. Communicating with Hedgehogs. *Nature Rev. Mol. Cell Biol.* **6**, 306–317 (2005).
- Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signalling: cell fate control and signal integration in development. *Science* **284**, 770–776 (1999).
- Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
- Taipale, J. & Beachy, P. A. The Hedgehog and Wnt signalling pathways in cancer. *Nature* **411**, 349–354 (2001).
- Allenspach, E. J., Maillard, I., Aster, J. C. & Pear, W. S. Notch signalling in cancer. *Cancer Biol. Ther.* **1**, 466–476 (2002).
- Lund, A. H. & van Lohuizen, M. Epigenetics and cancer. *Genes Dev.* **18**, 2315–2335 (2004).
- Valk-Lingbeek, M. E., Bruggeman, S. W. & van Lohuizen, M. Stem cells and cancer; the polycomb connection. *Cell* **118**, 409–418 (2004).
- Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38**, 413–443 (2004).
- Cao, R. & Zhang, Y. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr. Opin. Genet. Dev.* **14**, 155–164 (2004).
- Nguyen, C. T. et al. Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2'-deoxycytidine. *Cancer Res.* **62**, 6456–6461 (2002).
- Lachner, M., O'Sullivan, R. J. & Jenuwein, T. An epigenetic road map for histone lysine methylation. *J. Cell Sci.* **116**, 2117–2124 (2003).
- Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nature Rev. Cancer* **4**, 143–153 (2004).
- Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
- Woodhouse, E. C. & Liotta, L. A. *Drosophila* invasive tumors: a model for understanding metastasis. *Cell Cycle* **3**, 38–40 (2004).
- Dominguez, M. & Casares, F. Organ specification-growth control connection: new insights from the *Drosophila* eye-antennal disc. *Dev. Dyn.* **232**, 673–684 (2005).
- Toba, G. et al. The gene search system. A method for efficient detection and rapid molecular identification of genes in *Drosophila melanogaster*. *Genetics* **151**, 725–737 (1999).
- Dominguez, M. & de Celis, J. F. A dorsal/ventral boundary established by Notch controls growth and polarity in the *Drosophila* eye. *Nature* **396**, 276–278 (1998).
- Giniger, E., Tietje, K., Jan, L. Y. & Jan, Y. N. *lola* encodes a putative transcription factor required for axon growth and guidance in *Drosophila*. *Development* **120**, 1385–1398 (1994).
- Madden, K., Crowner, D. & Giniger, E. *LOLA* has the properties of a master regulator of axon-target interaction for SNb motor axons of *Drosophila*. *Dev. Biol.* **213**, 301–313 (1999).
- Goeke, S. et al. Alternative splicing of *lola* generates 19 transcription factors controlling axon guidance in *Drosophila*. *Nature Neurosci.* **6**, 917–924 (2003).
- Ohsako, T., Horiuchi, T., Matsuo, T., Komaya, S. & Aigaki, T. *Drosophila lola* encodes a family of BTB-transcription regulators with highly variable C-terminal domains containing zinc finger motifs. *Gene* **311**, 59–69 (2003).
- Horowitz, H. & Berg, C. A. The *Drosophila pipsqueak* gene encodes a nuclear BTB-domain-containing protein required early in oogenesis. *Development* **122**, 1859–1871 (1996).
- Weber, U., Siegel, V. & Mlodzik, M. *pipsqueak* encodes a novel nuclear protein required downstream of *seven-up* for the development of photoreceptors R3 and R4. *EMBO J.* **14**, 6247–6257 (1995).
- Siegmund, T. & Lehmann, M. The *Drosophila* Pipsqueak protein defines a new family of helix-turn-helix DNA-binding proteins. *Dev. Genes Evol.* **212**, 152–157 (2002).
- Siegel, V., Jongens, T. A., Jan, L. Y. & Jan, Y. N. *pipsqueak*, an early acting member of the posterior group of genes, affects vasa level and germ cell-somatic cell interaction in the developing egg chamber. *Development* **119**, 1187–1202 (1993).
- Lehmann, M., Siegmund, T., Lintermann, K. G. & Korge, G. The pipsqueak protein of *Drosophila melanogaster* binds to GAGA sequences through a novel DNA-binding domain. *J. Biol. Chem.* **273**, 28504–28509 (1998).
- Schwendemann, A. & Lehmann, M. Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. *Proc. Natl Acad. Sci. USA* **99**, 12883–12888 (2002).
- Rorth, P. et al. Systematic gain-of-function genetics in *Drosophila*. *Development* **125**, 1049–1057 (1998).
- Barna, M. et al. Plzf mediates transcriptional repression of HoxD gene expression through chromatin remodeling. *Dev. Cell* **3**, 499–510 (2002).
- Melnick, A. et al. Critical residues within the BTB domain of PLZF and Bcl-6 modulate interaction with corepressors. *Mol. Cell Biol.* **22**, 1804–1818 (2002).
- Schotta, G. et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).
- Irvine, K. D. Fringe, Notch, and making developmental boundaries. *Curr. Opin. Genet. Dev.* **9**, 434–441 (1999).
- Dominguez, M., Ferrer-Marco, D., Gutierrez-Avino, F. J., Speicher, S. A. & Benayahu, M. Growth and specification of the eye are controlled independently by Eyeless and Eyeless in *Drosophila melanogaster*. *Nature Genet.* **36**, 31–39 (2004).
- Chao, J. L., Tsai, Y. C., Chiu, S. J. & Sun, Y. H. Localized Notch signal acts through eyg and upd to promote global growth in *Drosophila* eye. *Development* **131**, 3839–3847 (2004).
- Bray, S., Musisi, H. & Bienz, M. Bre1 is required for Notch signalling and histone modification. *Dev. Cell* **8**, 279–286 (2005).
- Muller, J. et al. Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
- Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E. & Harte, P. J. The *Drosophila* Polycomb Group proteins ESC and EZ are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development* **128**, 275–286 (2001).

39. Czermin, B. *et al.* *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–196 (2002).
40. Janody, F. *et al.* A mosaic genetic screen reveals distinct roles for trithorax and polycomb group genes in *Drosophila* eye development. *Genetics* **166**, 187–200 (2004).
41. Harbour, J. W. & Dean, D. C. The Rb/E2F pathway: expanding roles and emerging paradigms. *Genes Dev.* **14**, 2393–2409 (2000).
42. Gowher, H., Leismann, O. & Jeltsch, A. DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO J.* **19**, 6918–6923 (2000).
43. Jabbari, K. & Bernardi, G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* **333**, 143–149 (2004).
44. Salzberg, A., Fisher, O., Siman-Tov, R. & Ankri, S. Identification of methylated sequences in genomic DNA of adult *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* **322**, 465–469 (2004).
45. Kunert, N., Marhold, J., Stanke, J., Stach, D. & Lyko, F. A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development* **130**, 5083–5090 (2003).
46. Hung, M. S. *et al.* *Drosophila* proteins related to vertebrate DNA (5-cytosine) methyltransferases. *Proc. Natl Acad. Sci. USA* **96**, 11940–11945 (1999).
47. Narsa Reddy, M., Tang, L. Y., Lee, T. L. & James Shen, C. K. A candidate gene for *Drosophila* genome methylation. *Oncogene* **22**, 6301–6303 (2003).
48. Marhold, J., Kramer, K., Kremmer, E. & Lyko, F. The *Drosophila* MBD2/3 protein mediates interactions between the MI-2 chromatin complex and CpT/A-methylated DNA. *Development* **131**, 6033–6039 (2004).
49. Baonza, A. & Freeman, M. Control of cell proliferation in the *Drosophila* eye by notch signalling. *Dev. Cell* **8**, 529–539 (2005).
50. Firth, L. C. & Baker, N. E. Extracellular signals responsible for spatially regulated proliferation in the differentiating *Drosophila* eye. *Dev. Cell* **8**, 541–551 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Müller, E. Giniger, A. Schwendemann, G. Reuter, F. Lyko, W. Chia, and A. Baonza for reagents; the Bloomington Stock Centre and Exelixis for fly stocks; the Developmental Studies Hybridoma Bank for antibodies; E. Ballesta-Illan for technical assistance; F. J. Garcia-Cozar for sharing quantitative RT-PCR expertise; L. A. Garcia-Alonso, J. Galceran and F. Viana for critically reading the manuscript; and M. Sefton for improvements to the manuscript. I.G.G. is a fellow of the CSIC I3P Programme. This work was supported by grants from the ‘Ministerio de Educación y Ciencia’ of Spain and a European Molecular Biology Organization Young Investigator Award to M.D.

Author Contributions D.F.-M. conceived the experiment to isolate the mutants; D.F.-M., I.G.-G., D.M.V., F.J.G.-A. and J.B. performed the experiments; and M.D. designed the experiments, carried out the data analysis and wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.D. (m.dominguez@umh.es).

Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing

J.-P. Beaulieu^{1,4}, D. P. Bennett^{1,3,5}, P. Fouqué^{1,6}, A. Williams^{1,7}, M. Dominik^{1,8}, U. G. Jørgensen^{1,9}, D. Kubas^{1,10}, A. Cassan^{1,4}, C. Coutures^{1,11}, J. Greenhill^{1,12}, K. Hill^{1,12}, J. Menzies^{1,13}, P. D. Sackett^{1,14}, M. Albrow^{1,15}, S. Brilliant^{1,10}, J. A. R. Caldwell^{1,16}, J. J. Calitz^{1,17}, K. H. Cook^{1,18}, E. Corrales^{1,4}, M. Desort^{1,4}, S. Dieters^{1,12}, D. Dominis^{1,19}, J. Donatowicz^{1,20}, M. Hoffman^{1,19}, S. Kane^{1,21}, J.-B. Marquette^{1,4}, R. Martin^{1,7}, P. Meintjes^{1,17}, K. Pollard^{1,15}, K. Sahu^{1,22}, C. Vinter^{1,9}, J. Wambsganss^{1,23}, K. Woller^{1,9}, K. Horne^{1,8}, I. Steele^{1,24}, D. M. Bramich^{1,8,24}, M. Burgdorf^{1,24}, C. Snodgrass^{1,25}, M. Bode^{1,24}, A. Udalski^{2,26}, M. K. Szymański^{2,26}, M. Kubiak^{2,26}, T. Więckowski^{2,26}, G. Pietrzyński^{2,26,27}, I. Soszyński^{2,26,27}, O. Szewczyk^{2,26}, Ł. Wyrzykowski^{2,26,28}, B. Paczyński^{2,29}, F. Abe^{3,30}, I. A. Bond^{3,31}, T. R. Britton^{3,15,32}, A. C. Gilmore^{3,15}, J. B. Hearnshaw^{3,15}, Y. Itow^{3,30}, K. Kamiya^{3,30}, P. M. Kilmartin^{3,15}, A. V. Korpela^{3,33}, K. Masuda^{3,30}, Y. Matsubara^{3,30}, M. Motomura^{3,30}, Y. Muraki^{3,30}, S. Nakamura^{3,30}, C. Okada^{3,30}, K. Ohnishi^{3,34}, N. J. Rattenbury^{3,28}, T. Sako^{3,30}, S. Sato^{3,35}, M. Sasaki^{3,30}, T. Sekiguchi^{3,30}, D. J. Sullivan^{3,33}, P. J. Tristram^{3,32}, P. C. M. Yock^{3,32} & T. Yoshioka^{3,30}

In the favoured core-accretion model of formation of planetary systems, solid planetesimals accumulate to build up planetary cores, which then accrete nebular gas if they are sufficiently massive. Around M-dwarf stars (the most common stars in our Galaxy), this model favours the formation of Earth-mass ($M_{\oplus}$) to Neptune-mass planets with orbital radii of 1 to 10 astronomical units (AU), which is consistent with the small number of gas giant planets known to orbit M-dwarf host stars^{1–4}. More than 170 extrasolar planets have been discovered with a wide range of masses and orbital periods, but planets of Neptune's mass or less have not hitherto been detected at separations of more than 0.15 AU from normal stars. Here we report the discovery of a $5.5_{-2.7}^{+5.5} M_{\oplus}$ planetary companion at a separation of $2.6_{-0.6}^{+1.5}$ AU from a $0.22_{-0.11}^{+0.21} M_{\odot}$ M-dwarf star, where $M_{\odot}$ refers to a solar mass. (We propose to name it OGLE-2005-BLG-390Lb, indicating a planetary mass companion to the lens star of the microlensing event.) The mass is lower than that of GJ876d (ref. 5), although the error bars overlap. Our detection suggests that such cool, sub-Neptune-mass planets may be more common than gas giant planets, as predicted by the core accretion theory.

Gravitational microlensing events can reveal extrasolar planets orbiting the foreground lens stars if the light curves are measured frequently enough to characterize planetary light curve deviations with features lasting a few hours^{6–9}. Microlensing is most sensitive to planets in Earth-to-Jupiter-like orbits with semi-major axes in the range 1–5 AU. The sensitivity of the microlensing method to low-mass planets is restricted by the finite angular size of the source stars^{10,11}, limiting detections to planets of a few $M_{\oplus}$ for giant source stars, but allowing the detection of planets as small as $0.1 M_{\oplus}$ for main-sequence source stars in the Galactic Bulge. The PLANET collaboration¹² maintains the high sampling rate required to detect low-mass planets while monitoring the most promising of the >500 microlensing events discovered annually by the OGLE collaboration, as well as events discovered by MOA. A decade of pioneering microlensing searches has resulted in the recent detections of two Jupiter-mass extrasolar planets^{13,14} with orbital separations of a few AU by the combined observations of the OGLE, MOA, MicroFUN and PLANET collaborations. The absence of perturbations to stellar microlensing events can be used to constrain the presence of planetary lens companions. With large samples of events, upper

¹PLANET/RoboNet Collaboration (<http://planet.iap.fr> and <http://www.astro.livjm.ac.uk/RoboNet/>). ²OGLE Collaboration (<http://ogle.astrouw.edu.pl>). ³MOA Collaboration (<http://www.physics.auckland.ac.nz/moa>). ⁴Institut d'Astrophysique de Paris, CNRS, Université Pierre et Marie Curie UMR7095, 98bis Boulevard Arago, 75014 Paris, France. ⁵University of Notre Dame, Department of Physics, Notre Dame, Indiana 46556-5670, USA. ⁶Observatoire Midi-Pyrénées, Laboratoire d'Astrophysique, UMR 5572, Université Paul Sabatier—Toulouse 3, 14 avenue Edouard Belin, 31400 Toulouse, France. ⁷Perth Observatory, Walnut Road, Bickley, Perth, WA 6076, Australia. ⁸Scottish Universities Physics Alliance, University of St Andrews, School of Physics and Astronomy, North Haugh, St Andrews KY16 9SS, UK. ⁹Niels Bohr Institutet, Astronomisk Observatorium, Juliane Maries Vej 30, 2100 København Ø, Denmark. ¹⁰European Southern Observatory, Casilla 19001, Santiago 19, Chile. ¹¹CEA DAPNIA/SPP Saclay, 91191 Gif-sur-Yvette cedex, France. ¹²University of Tasmania, School of Mathematics and Physics, Private Bag 37, Hobart, TAS 7001, Australia. ¹³South African Astronomical Observatory, PO Box 9, Observatory 7935, South Africa. ¹⁴Research School of Astronomy and Astrophysics, Australian National University, Mt Stromlo Observatory, Weston Creek, ACT 2611, Australia. ¹⁵University of Canterbury, Department of Physics and Astronomy, Private Bag 4800, Christchurch 8020, New Zealand. ¹⁶McDonald Observatory, 16120 St Hwy Spur 78 #2, Fort Davis, Texas 79734, USA. ¹⁷Boyden Observatory, University of the Free State, Department of Physics, PO Box 339, Bloemfontein 9300, South Africa. ¹⁸Lawrence Livermore National Laboratory, IGGP, PO Box 808, Livermore, California 94551, USA. ¹⁹Universität Potsdam, Institut für Physik, Am Neuen Palais 10, 14469 Potsdam, Astrophysikalisches Institut Potsdam, An der Sternwarte 16, D-14482, Potsdam, Germany. ²⁰Technische Universität Wien, Wiedner Hauptstrasse 8 / 020 B.A. 1040 Wien, Austria. ²¹Department of Astronomy, University of Florida, 211 Bryant Space Science Center, Gainesville, Florida 32611-2055, USA. ²²Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ²³Astronomisches Rechen-Institut (ARI), Zentrum für Astronomie, Universität Heidelberg, Mönchhofstrasse 12-14, 69120 Heidelberg, Germany. ²⁴Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead CH41 1LD, UK. ²⁵Astronomy and Planetary Science Division, Department of Physics, Queen's University Belfast, Belfast, UK. ²⁶Observatorium Astronomiczne Uniwersytetu Warszawskiego, Aleje Ujazdowskie 4, 00-478 Warszawa, Poland. ²⁷Universidad de Concepcion, Departamento de Física, Casilla 160-C, Concepcion, Chile. ²⁸Jodrell Bank Observatory, The University of Manchester, Macclesfield, Cheshire SK11 9DL, UK. ²⁹Princeton University Observatory, Peyton Hall, Princeton, New Jersey 08544, USA. ³⁰Solar-Terrestrial Environment Laboratory, Nagoya University, Nagoya 464-860, Japan. ³¹Institute for Information and Mathematical Sciences, Massey University, Private Bag 102-904, Auckland, New Zealand. ³²Department of Physics, University of Auckland, Private Bag 92019, Auckland, New Zealand. ³³School of Chemical and Physical Sciences, Victoria University, PO Box 600, Wellington, New Zealand. ³⁴Nagano National College of Technology, Nagano 381-8550, Japan. ³⁵Department of Astrophysics, Faculty of Science, Nagoya University, Nagoya 464-860, Japan.

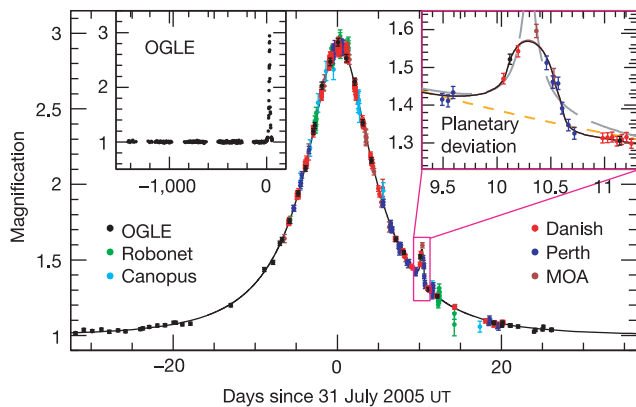


Figure 1 | The observed light curve of the OGLE-2005-BLG-390 microlensing event and best-fit model plotted as a function of time. Error bars are 1σ . The data set consists of 650 data points from PLANET Danish (ESO La Silla, red points), PLANET Perth (blue), PLANET Canopus (Hobart, cyan), RoboNet Faulkes North (Hawaii, green), OGLE (Las Campanas, black), MOA (Mt John Observatory, brown). This photometric monitoring was done in the I band (with the exception of the Faulkes R-band data and the MOA custom red passband) and real-time data reduction was performed with the different OGLE, PLANET and MOA data reduction pipelines. Danish and Perth data were finally reduced by the image subtraction technique¹⁹ with the OGLE pipeline. The top left inset shows the OGLE light curve extending over the previous 4 years, whereas the top right one shows a zoom of the planetary deviation, covering a time interval of 1.5 days. The solid curve is the best binary lens model described in the text with $q = 7.6 \pm 0.7 \times 10^{-5}$, and a projected separation of $d = 1.610 \pm 0.008 R_E$. The dashed grey curve is the best binary source model that is rejected by the data, and the dashed orange line is the best single lens model.

limits on the frequency of Jupiter-mass planets have been placed over an orbital range of 1–10 AU, down to $M_{\oplus}$ planets^{15–17} for the most common stars of our galaxy.

On 11 July 2005, the OGLE Early Warning System¹⁸ announced the microlensing event OGLE-2005-BLG-390 (right ascension $\alpha = 17^{\text{h}} 54^{\text{m}} 19.2^{\text{s}}$, declination $\delta = -30^{\circ} 22' 38''$, J2000) with a relatively bright clump giant as a source star. Subsequently, PLANET, OGLE and MOA monitored it with their different telescopes. After peaking at a maximum magnification of $A_{\text{max}} = 3.0$ on 31 July 2005, a short-duration deviation from a single lens light curve was detected on 9 August 2005 by PLANET. As described below, this deviation was due to a low-mass planet orbiting the lens star.

From analysis of colour-magnitude diagrams, we derive the following reddening-corrected colours and magnitudes for the source star: $(V - I)_0 = 0.85$, $I_0 = 14.25$ and $(V - K)_0 = 1.9$. We used the surface brightness relation²⁰ linking the emerging flux per solid angle of a light-emitting body to its colour, calibrated by interferometric observations, to derive an angular radius of $5.25 \pm 0.73 \mu\text{as}$, which corresponds to a source radius of $9.6 \pm 1.3 R_{\odot}$ (where $R_{\odot}$ is the radius of the Sun) if the source star is at a distance of 8.5 kpc. The source star colours indicate that it is a 5,200 K giant, which corresponds to a G4 III spectral type.

Figure 1 shows our photometric data for microlensing event OGLE-2005-BLG-390 and the best planetary binary lens model. The best-fit model has $\chi^2 = 562.26$ for 650 data points, seven lens parameters, and 12 flux normalization parameters, for a total of 631 degrees of freedom. Model length parameters in Table 1 are expressed in units of the Einstein ring radius R_E (typically ~ 2 AU for a Galactic Bulge system), the size of the ring image that would be seen in the case of perfect lens–source alignment. In modelling the light curve, we adopted linear limb darkening laws²¹ with $\Gamma_I = 0.538$ and $\Gamma_R = 0.626$, appropriate for this G4 III giant source star, to describe

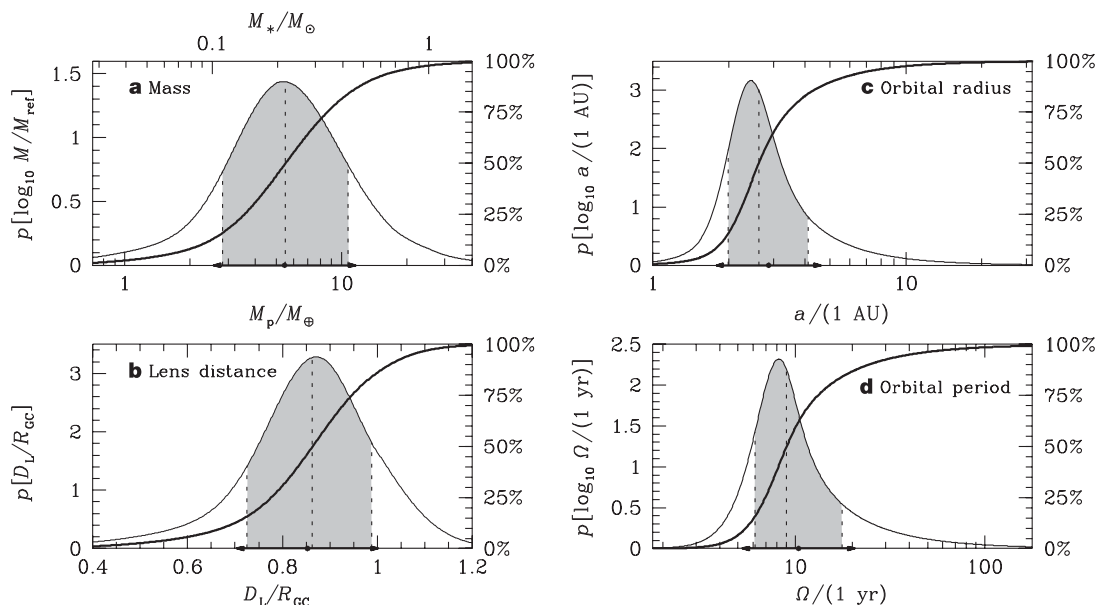


Figure 2 | Bayesian probability densities for the properties of the planet and its host star. **a**, The masses of the lens star and its planet (M_* and M_p respectively), **b**, their distance from the observer (D_L), **c**, the three-dimensional separation or semi-major axis a of an assumed circular planetary orbit; and **d**, the orbital period Ω of the planet. (In **a**, M_{ref} refers to $M_{\oplus}$ on the upper x axis and $M_{\odot}$ on the lower x axis.) The bold, curved line in each panel is the cumulative distribution, with the percentiles listed on the right. The dashed vertical lines indicate the medians, and the shading indicates the central 68.3% confidence intervals, while dots and arrows on the abscissa mark the expectation value and standard deviation. All estimates follow from a Bayesian analysis assuming a standard model for the disk and bulge population of the Milky Way, the stellar mass function of

ref. 23, and a gaussian prior distribution for $D_S = 1.05 \pm 0.25 R_{GC}$ (where $R_{GC} = 7.62 \pm 0.32$ kpc for the Galactic Centre distance). The medians of these distributions yield a $5.5^{+5.5}_{-2.7} M_{\oplus}$ planetary companion at a separation of $2.6^{+1.5}_{-0.6}$ AU from a $0.22^{+0.21}_{-0.11} M_{\odot}$ Galactic Bulge M-dwarf at a distance of 6.6 ± 1.0 kpc from the Sun. The median planetary period is 9^{+9}_{-3} years. The logarithmic means of these probability distributions (which obey Kepler's third law) are a separation of 2.9 AU, a period of 10.4 years, and masses of $0.22 M_{\odot}$ and $5.5 M_{\oplus}$ for the star and planet, respectively. In each plot, the independent variable for the probability density is listed within square brackets. The distribution of the planet–star mass ratio was taken to be independent of the stellar mass, and a uniform prior distribution was assumed for the planet–star separation distribution.

Table 1 | Microlensing fit parameters

d	$1.610 \pm 0.008 R_E$
q	$(7.6 \pm 0.7) \times 10^{-5}$
Closest approach	$0.359 \pm 0.005 R_E$
Einstein ring radius crossing time	11.03 ± 0.11 days
Time of closest approach	31.231 ± 0.005 July 2005 UT
Source star radius crossing time	0.282 ± 0.010 days
θ	2.756 ± 0.003 rad

The parameters for the best binary lens model for the OGLE 2005-BLG-390 microlensing event light curve are shown with their 1σ uncertainties. Some of these parameters are scaled to the Einstein ring radius, which is given by $R_E = 2\sqrt{GMD_L(D_S - D_L)/(c^2 D_S)}$, where M is the mass of the lens, G is the newtonian constant of gravitation, c is the speed of light in vacuum, and D_L and D_S are the lens and source distances, respectively.

the centre-to-limb variation of the intensity profile in the I and R bands. Four different binary lens modelling codes were used to confirm that the model we present is the only acceptable model for the observed light curve. The best alternative model is one with a large-flux-ratio binary source with a single lens, which has gross features that are similar to a planetary microlensing event²². However, as shown in Fig. 1, this model fails to account for the PLANET-Perth, PLANET-Danish and OGLE measurements near the end of the planetary deviation, and it is formally excluded by $\Delta\chi^2 = 46.25$ with one less model parameter.

The planet is designated OGLE-2005-BLG-390Lb, where the 'Lb' suffix indicates the secondary component of the lens system with a planetary mass ratio. The microlensing fit only directly determines the planet–star mass ratio, $q = 7.6 \pm 0.7 \times 10^{-5}$, and the projected planet–star separation, $d = 1.610 \pm 0.008 R_E$. Although the planet and star masses are not directly determined for planetary microlensing events, we can derive their probability densities. We have performed a bayesian analysis²³ employing the Galactic models and mass functions described in refs 11 and 23. We averaged over the distances and velocities of the lens and source stars, subject to the constraints due to the angular diameter of the source and the measured parameters given in Table 1. This analysis gives a 95% probability that the planetary host star is a main-sequence star, a 4% probability that it is a white dwarf, and a probability of <1% that it is a neutron star or black hole. The host star and planet parameter probability densities for a main sequence lens star are shown in Fig. 2 for the Galactic model used in ref. 23. The medians of the lens parameter probability distributions yield a companion mass of $5.5^{+5.5}_{-2.7} M_{\oplus}$ and an orbital separation of $2.6^{+1.5}_{-0.6}$ AU from the $0.22^{+0.21}_{-0.11} M_{\odot}$ lens star, which is located at a distance of $D_L = 6.6 \pm 1.0$ kpc. These error bars indicate the central 68% confidence interval. These median parameters imply that the planet receives radiation from its host star that is only 0.1% of the radiation that the Earth receives from the Sun, so the probable surface temperature of the planet is ~ 50 K, similar to the temperatures of Neptune and Pluto.

The parameters of this event are near the limits of microlensing planet detectability for a giant source star. The separation of $d = 1.61$ is near the outer edge of the so-called lensing zone⁷, which has the highest planet detection probability, and the planet's mass is about a factor of two above the detection limit set by the finite size of the source star. Planets with $q > 10^{-3}$ and $d \approx 1$ are much easier to detect, and so it may be that the parameters of OGLE-2005-BLG-390Lb represent a more common type of planet. This can be quantified by simulating planetary light curves with different values of q and θ (where θ is the angle of source motion with respect to the lens axis) but the remaining parameters are fixed to the values for the three known microlensing planets. We find that the probability of detecting a $q \approx 4\text{--}7 \times 10^{-3}$ planet, like the first two microlens planets^{13,14}, is ~ 50 times larger than the probability of detecting a $q = 7.6 \times 10^{-5}$ planet like OGLE-2005-BLG-390Lb. This suggests that, at the orbital separations probed by microlensing, sub-Neptune-mass planets are significantly more common

than large gas giants around the most common stars in our Galaxy. Similarly, the first detection of a sub-Neptune-mass planet at the outer edge of the 'lensing zone' provides a hint that these sub-Neptune-mass planets may tend to reside in orbits with semi-major axes $a > 2$ AU.

The core-accretion model of planet formation predicts that rocky/icy $5\text{--}15 M_{\oplus}$ planets orbiting their host stars at $1\text{--}10$ AU are much more common than Jupiter-mass planets, and this prediction is consistent with the small fraction of M-dwarfs with planets detected by radial velocities^{3,5} and with previous limits from microlensing¹⁵. Our discovery of such a low-mass planet by gravitational microlensing lends further support to this model, but more detections of similar and lower-mass planets over a wide range of orbits are clearly needed. Planets with separations of ~ 0.1 AU will be detected routinely by the radial velocity method or space observations of planetary transits in the coming years^{24–27}, but the best chance to increase our understanding of such planets over orbits of $1\text{--}10$ AU in the next $5\text{--}10$ years is by future interferometer programs²⁸ and more advanced microlensing surveys^{11,29,30}.

Received 28 September; accepted 14 November 2005.

1. Safronov, V. *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets* (Nauka, Moscow, 1969).
2. Wetherill, G. W. Formation of the terrestrial planets. *Annu. Rev. Astron. Astrophys.* **18**, 77–113 (1980).
3. Laughlin, G., Bodenheimer, P. & Adams, F. C. The core accretion model predicts few jovian-mass planets orbiting red dwarfs. *Astrophys. J.* **612**, L73–L76 (2004).
4. Ida, S. & Lin, D. N. C. Toward a deterministic model of planetary formation. II. The formation and retention of gas giant planets around stars with a range of metallicities. *Astrophys. J.* **616**, 567–572 (2004).
5. Rivera, E. et al. A ~ 7.5 Earth-mass planet orbiting the nearby star, GJ 876. *Astrophys. J.* (in the press).
6. Mao, S. & Paczynski, B. Gravitational microlensing by double stars and planetary systems. *Astrophys. J.* **374**, L37–L40 (1991).
7. Gould, A. & Loeb, A. Discovering planetary systems through gravitational microlenses. *Astrophys. J.* **396**, 104–114 (1992).
8. Wambsganss, J. Discovering Galactic planets by gravitational microlensing: magnification patterns and light curves. *Mon. Not. R. Astron. Soc.* **284**, 172–188 (1997).
9. Griest, K. & Safizadeh, N. The use of high-magnification microlensing events in discovering extrasolar planets. *Astrophys. J.* **500**, 37–50 (1998).
10. Bennett, D. P. & Rhie, S. H. Detecting Earth-mass planets with gravitational microlensing. *Astrophys. J.* **472**, 660–664 (1996).
11. Bennett, D. P. & Rhie, S. H. Simulation of a space-based microlensing survey for terrestrial extrasolar planets. *Astrophys. J.* **574**, 985–1003 (2002).
12. Albrow, M. et al. The 1995 pilot campaign of PLANET: searching for microlensing anomalies through precise, rapid, round-the-clock monitoring. *Astrophys. J.* **509**, 687–702 (1998).
13. Bond, I. A. et al. OGLE 2003-BLG-235/MOA 2003-BLG-53: A planetary microlensing event. *Astrophys. J.* **606**, L155–L158 (2004).
14. Udalski, A. et al. A jovian-mass planet in microlensing event OGLE-2005-BLG-071. *Astrophys. J.* **628**, L109–L112 (2005).
15. Gaudi, B. S. et al. Microlensing constraints on the frequency of Jupiter-mass companions: analysis of 5 years of PLANET photometry. *Astrophys. J.* **566**, 463–499 (2002).
16. Abe, F. et al. Search for low-mass exoplanets by gravitational microlensing at high magnification. *Science* **305**, 1264–1267 (2004).
17. Dong, S. et al. Planetary detection efficiency of the magnification 3000 microlensing event OGLE-2004-BLG-343. *Astrophys. J.* (submitted); preprint at (<http://arXiv.org/astro-ph/0507079>) (2005).
18. Udalski, A. The optical gravitational lensing experiment. real time data analysis systems in the OGLE-III survey. *Acta Astron.* **53**, 291–305 (2003).
19. Alard, C. Image subtraction using a space-varying kernel. *Astron. Astrophys. Suppl.* **144**, 363–370 (2000).
20. Kervella, P. et al. Cepheid distances from infrared long-baseline interferometry. III. Calibration of the surface brightness-color relations. *Astron. Astrophys.* **428**, 587–593 (2004).
21. Claret, A., Diaz-Cordoves, J. & Gimenez, A. Linear and non-linear limb-darkening coefficients for the photometric bands R I J H K. *Astron. Astrophys. Suppl.* **114**, 247–252 (1995).
22. Gaudi, B. S. Distinguishing between binary-source and planetary microlensing perturbations. *Astrophys. J.* **506**, 533–539 (1998).
23. Dominik, M. Stochastic distributions of lens and source properties for observed galactic microlensing events. *Mon. Not. R. Astron. Soc.* (submitted); preprint at (<http://arXiv.org/astro-ph/0507540>) (2005).

24. Vogt, S. S. *et al.* Five new multicomponent planetary systems. *Astrophys. J.* **632**, 638–658 (2005).
25. Mayor, M. *et al.* The CORALIE survey for southern extrasolar planets. XII. Orbital solutions for 16 extrasolar planets discovered with CORALIE. *Astron. Astrophys.* **415**, 391–402 (2004).
26. Borucki, W. *et al.* in *Second Eddington Workshop: Stellar Structure and Habitable Planet Finding* (eds Favata, F., Aigrain, S. & Wilson, A.) 177–182 (ESA SP-538, ESA Publications Division, Noordwijk, 2004).
27. Moutou, C. *et al.* Comparative blind test of five planetary transit detection algorithms on realistic synthetic light curves. *Astron. Astrophys.* **437**, 355–368 (2005).
28. Sozzetti, A. *et al.* Narrow-angle astrometry with the space interferometry mission: the search for extrasolar planets. I. Detection and characterization of single planets. *Pub. Astron. Soc. Pacif.* **114**, 1173–1196 (2002).
29. Bennett, D. P. in *ASP Conf. Ser. on Extrasolar Planets: Today and Tomorrow* (eds Beaulieu, J.-P., Lecavelier des Etangs, A. & Terquem, C.) Vol. 321, 59–68 (ASP, 2004).
30. Beaulieu, J. P. *et al.* PLANET III: searching for Earth-mass planets via microlensing from Dome C? *ESA Publ. Ser.* **14**, 297–302 (2005).

Acknowledgements PLANET is grateful to the observatories that support our science (the European Southern Observatory, Canopus, Perth; and the South African Astronomical Observatory, Boyden, Faulkes North) and to the ESO team in La Silla for their help in maintaining and operating the Danish telescope. Support for the PLANET project was provided by CNRS, NASA, the NSF, the LLNL/NNSA/DOE, PNP, PICS France-Australia, D. Warren, the DFG, IDA and the SNF. RoboNet is funded by the UK PPARC and the FTN was supported by the Dill Faulkes Educational Trust. Support for the OGLE project, conducted at Las Campanas Observatory (operated by the Carnegie Institution of Washington), was provided by the Polish Ministry of Science, the Foundation for Polish Science, the NSF and NASA. The MOA collaboration is supported by MEXT and JSPS of Japan, and the Marsden Fund of New Zealand.

Author Information The photometric data set is available at planet.iap.fr and ogle.astrouw.edu.pl. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.P.B. (beaulieu@iap.fr) or D.P.B. (bennett@nd.edu).

Laser acceleration of quasi-monoenergetic MeV ion beams

B. M. Hegelich¹, B. J. Albright¹, J. Cobble¹, K. Flippo¹, S. Letzring¹, M. Paffett¹, H. Ruhl², J. Schreiber^{3,4}, R. K. Schulze¹ & J. C. Fernández¹

Acceleration of particles by intense laser–plasma interactions represents a rapidly evolving field of interest, as highlighted by the recent demonstration^{1–4} of laser-driven relativistic beams of monoenergetic electrons. Ultrahigh-intensity lasers can produce accelerating fields of 10 TV m^{-1} ($1 \text{ TV} = 10^{12} \text{ V}$), surpassing those in conventional accelerators by six orders of magnitude. Laser-driven ions with energies of several MeV per nucleon have also been produced^{5–9}. Such ion beams exhibit unprecedented characteristics—short pulse lengths, high currents and low transverse emittance¹⁰—but their exponential energy spectra have almost 100% energy spread. This large energy spread, which is a consequence of the experimental conditions used to date, remains the biggest impediment to the wider use of this technology. Here we report the production of quasi-monoenergetic laser-driven C^{5+} ions with a vastly reduced energy spread of 17%. The ions have a mean energy of 3 MeV per nucleon (full-width at half-maximum $\sim 0.5 \text{ MeV}$ per nucleon) and a longitudinal emittance of less than $2 \times 10^{-6} \text{ eV s}$ for pulse durations shorter than 1 ps. Such laser-driven, high-current, quasi-monoenergetic ion sources may enable significant advances in the development of compact MeV ion accelerators¹¹, new diagnostics^{12,13}, medical physics¹⁴, inertial confinement fusion and fast ignition^{15–17}.

An ultrahigh-intensity laser ($I\lambda^2 > 10^{18} \text{ W cm}^{-2} \mu\text{m}^{-2}$, where I is intensity and λ is wavelength) incident on a target accelerates a large number of electrons to multi-MeV energies^{18,19}. These electrons traverse typical thin foil targets and set up a very strong electrostatic field exceeding 1 TV m^{-1} . This field ionizes the rear surface and accelerates ions to energies of many MeV. This process is known as target normal sheath acceleration (TNSA)⁵. Experiments have demonstrated acceleration of protons to more than 60 MeV (ref. 8), fluorine ions to above 100 MeV (ref. 6) and high- Z palladium ions up to 225 MeV (ref. 20), that is, more than 2 MeV per nucleon. These ion beams have a much lower transverse temperature and a much shorter duration and a much higher current than those from conventional accelerators. These unique characteristics make them ideal candidates for a number of experiments not feasible otherwise.

Owing to their short pulse length and high energy content, the ion beams can heat macroscopic amounts of matter to more than 10^6 °C before the matter can expand²¹, thereby creating conditions of high temperature and density only found in the interior of stars. Conversely they can also be used as a probe to investigate ion transport and stopping in a hot, dense plasma before it has time to disassemble. Conventional accelerators are hard pressed to deliver enough particles in the available $\sim \text{ps}$ time window to make high-quality measurements feasible. These are but two examples where the high current and short pulse duration are the key to an otherwise impossible experiment. More examples can be found in nuclear

physics, fusion research and other areas—examples are the synthesis of neutron rich nuclei or the measurement of fusion cross-sections in supernova-like hot, dense plasma conditions. The much higher beam current and the much lower emittance of the laser-driven ion beams make them a promising candidate for advanced accelerator concepts.

Today, a standard linear accelerator that matches the MeV/u energy level (with u being the atomic mass unit) of these laser-driven ions is $\sim 100 \text{ m}$ long. In contrast, the laser fits in a large room and accelerates the ions to MeV/u over just $10 \mu\text{m}$. The low duty-cycle in present experiments is a limitation that is likely to be mitigated by the next generation of high-power lasers, currently under development. However, the major difficulty with all the TNSA and other laser-driven ion-acceleration mechanisms^{9,22} has been the resulting maxwellian energy distribution, with a typical 100% energy spread^{6,8,9}. All the above-mentioned applications would benefit greatly from a narrower energy distribution, centred about a specific value.

We report here a laser-driven quasi-monoenergetic ion beam, a

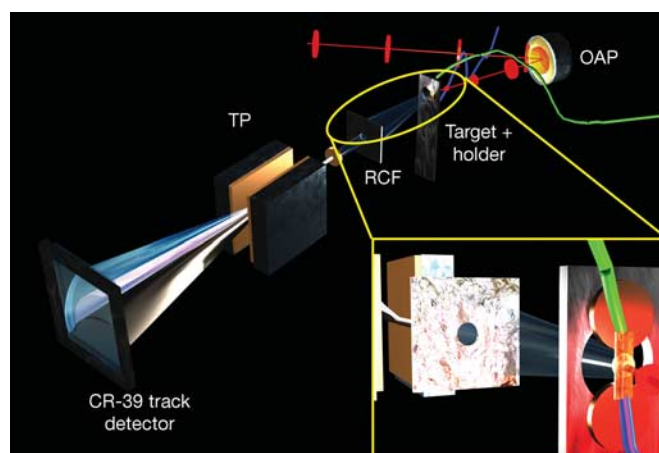


Figure 1 | Experimental set-up. A short, high-intensity laser pulse is focused on a thin metal foil target by an off-axis parabolic mirror (OAP). The red line shows the laser beam axis, and the red disks represent the laser pulse travelling along that axis and getting focused down by the OAP. Two wires (green and blue) are attached to the target, pass a current through it and heat it to $\sim 1,100 \text{ K}$ to remove contaminants. Ions are accelerated at the target rear surface and are detected by a stack of radiochromic film (RCF) and a Thomson parabola (TP) spectrometer using CR-39 track detectors. The inset shows an enlarged frontside view of the target, with the target and the green and blue wires being in the lower right corner and the RCF and TP detectors in the upper left.

¹Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ²University of Nevada, Reno, Nevada 89557, USA. ³Ludwig-Maximilians-Universität München, München, Germany.

⁴Max-Planck-Institut für Quantenoptik, Garching 85748, Germany.

C^{5+} beam created in the interaction of a 20 TW/0.8 ps laser pulse with a solid target. A simple schematic illustrating the process can be found in Supplementary Information (sections SI_1 and SI_2). The experiments were performed at the LANL Trident laser facility. The experimental set-up is shown in Fig. 1 (for details see Methods), and a time-integrated photograph of an actual laser shot is shown in Supplementary Information section SI_3. The monoenergetic signature is the direct result of a fundamentally different target composition employed in these experiments. In TNSA, the ions with the highest charge-to-mass ratio dominate the acceleration, gaining the most energy. Given typical vacuum conditions of $\sim 10^{-6}$ mbar, surface target contaminants containing protons are always present. These protons have the largest charge-to-mass ratio by at least a factor of 2. Controlled treatment of foil targets before irradiation with the ultrahigh-intensity laser reduces adsorbed and absorbed proton contaminants to an unobservable level, allowing higher- Z ions to be the dominant species⁶. Using the right treatment parameters and target materials, a thin source layer of just a few monolayers can be formed by catalytic processes.

Specifically, we have demonstrated the acceleration of C^{5+} and C^{6+} from an ultrathin layer of graphitic carbon, formed from catalytic decomposition of adsorbed hydrocarbon impurities on a 20 μm palladium foil. Unlike the low-energy lasers which are used for electron acceleration^{2–4}, which have a high repetition rate and allow the taking of many shots to obtain good statistics, ion acceleration requires higher energy lasers which are single shot in nature. The number of shots is extremely limited and fluctuations in the laser parameters further complicate obtaining good statistics. However, five shots exhibiting monoenergetic carbon ions have been observed in two separate campaigns months apart, and another ten shots showing indication of monoenergetic ions are still being analysed. Figure 2 shows the measured C^{5+} spectrum (black curve) with the lowest ratio $\Delta E/E$ of $\sim 17\%$, where E is the mean energy of the C^{5+} ions and ΔE is their energy spread. It also shows the corresponding highest substrate charge state Pd^{22+} (blue). Having the highest

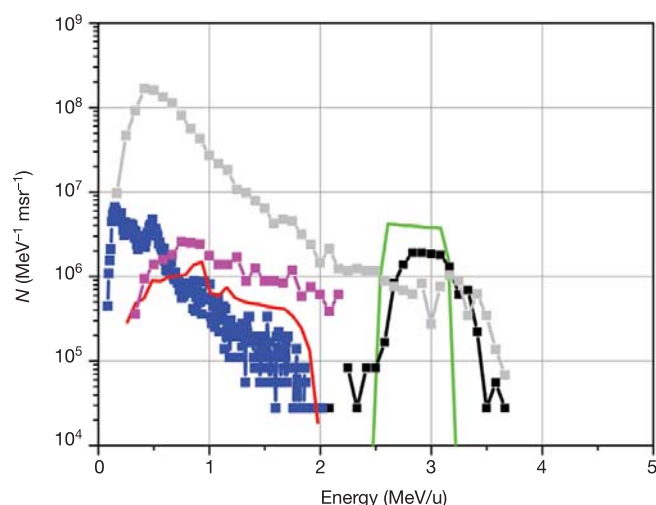


Figure 2 | Monoenergetic carbon ions from a 20 μm palladium substrate. The curves show ion number (N) over energy per nucleon (MeV/u). The black curve shows the spectra of the measured C^{5+} ions, the blue curve shows the dominant substrate charge state Pd^{22+} . The green and the red curves are simulations obtained using the 1D-hybrid-code BILBO, showing the simulated C^{5+} and Pd^{21+} spectra, respectively. The grey curve shows the dominant C^{4+} signal from a heated W target, and the magenta trace shows the C^{5+} signal from a cold Pd target. In these last two cases, the targets have a thick layer of carbon contaminants and do not form a monolayer source. The resulting carbon signals are therefore exponential and show lower numbers in the high-energy range. The errors are: $dN \leq 1\%$ statistical accuracy, and $dE \leq 2\%$ for C and $dE \leq 4.5\%$ for Pd.

charge-to-mass ratio of 0.42, the C^{5+} is dominantly accelerated. Owing to the extremely small spatial extent of the carbon layer and its localization at the rear surface, all of the carbon ions are accelerated at once at the peak of the accelerating field, leading to the monoenergetic ion pulse. After all carbon ions are accelerated, the field is still very strong and only moderately screened by the carbon, therefore the next highest charge-to-mass ratio ion—that is, Pd^{22+} with a charge-to-mass ratio of 0.2—is now dominantly accelerated and gains a large fraction of the energy before the field decays and lower Pd charge states are created and accelerated.

For the purpose of this Letter we limit our discussion to the two dominant charge states, which together contain $\sim 20\%$ of the total integrated ion energy and have a bearing on the results reported here. The leading short bunch of C^{5+} ions shows a monoenergetic energy distribution with a mean energy of $E \approx 36$ MeV, that is, 3 MeV per nucleon and a full-width at half-maximum of 0.5 MeV per nucleon. We infer that the accelerated C^{5+} ion bunch has a longitudinal emittance of $\epsilon_1 < 2 \times 10^{-6} \pi$ eV s, improving on conventional high-current accelerators by orders of magnitude. Also, in contrast to the Pd and to any previous measurements, no lower C charge states are present. Closer analysis reveals important differences in the acceleration mechanism for the Pd substrate ions and the C ions from the source surface layer. Whereas the substrate ions have a typical exponential spectrum, the C ions are monoenergetic.

The small energy spread of the observed carbon ions can be understood from consideration of quasi-neutral ($n_e = Z_{Pd}n_{Pd}$), adiabatic expansion in one dimension (1D) of a palladium substrate coated by a very thin film of carbon. (Here n_e is the electron density, n_{Pd} the palladium density and Z_{Pd} the mean palladium charge state.) The electric field obeys $eE \approx -m_e n_e^{-1} \partial_x \int dv v^2 (f_e - Z_{Pd} f_{Pd})$, with f_e and f_{Pd} the distribution functions of electrons and palladium ions, e the elementary charge, m_e the electron mass and v velocity. Such a plasma column will expand with sound speed c_s to characteristic size $L^2(t) = L_0^2(1 + t^2 c_s^2 / L_0^2)$. The ion and electron temperatures will therefore decrease by a factor $L_0/L(t)$, which leads to an electric field $eE \approx x c_s^2 m_{Pd} Z_{Pd}^{-2} L^{-2}$, where m_{Pd} is the atomic mass of palladium. This field leads to an acceleration $d^2 x_C / dt^2 = r x_C c_s^2 L^{-2}$ of the carbon ions, with x_C being the spatial coordinate of the carbon ions. The dynamics of the layer are characterized by r , the ratio of charge-to-mass ratios of C to Pd ions: for $r \gg 1$, the carbon layer detaches from the substrate at early time and propagates ahead of it as a directed bunch. For $r < 1$, the substrate overtakes the C layer and flow instabilities may arise. With an average Pd charge state $Z_{eff,Pd} \approx 7$, one obtains $r = 6.3$, predicting a clean separation of

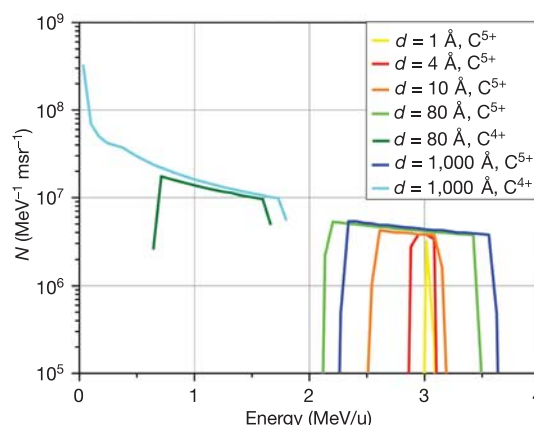


Figure 3 | Changing the thickness of the carbon source layer leads to a change in the energy spectrum in the BILBO simulations. Decreasing the layer thickness (d) causes the spectrum to become more monoenergetic. Increasing the layer thickness leads to a broader distribution and ultimately the appearance of lower charge states and a maxwellian spectrum.

the carbon ions from the substrate. Esirkepov *et al.*²³ have also examined the problem of monoenergetic ion acceleration, but their model relies explicitly upon finite transverse extent of the target and is not applicable to the TNSA scenario we find in our experiments.

In order to improve our understanding and our predictive capability, we developed a numerical model that simulates the ionization and acceleration physics. Full *ab initio* simulations with the required dynamics and sufficiently low noise levels to faithfully capture the ionization kinetics are not feasible, so we have focused on a reduced model that takes into account the essential physics. This 1D-hybrid model BILBO (backside ion lagrangian blow-off) uses a relativistic Boltzmann fluid model of the electrons and represents ions as kinetic simulation particles. This model has been explicitly designed to implement TNSA⁵ in a heterogeneous mixture of ionization species and ion types. In our simulations, a thin layer of carbon (1–1,000 Å) with areal density $\rho = 5 \times 10^{-10} - 5 \times 10^{-7} \text{ g cm}^{-2}$ is placed on the surface of a palladium foil of solid density $\rho = 12.16 \text{ g cm}^{-3}$ and a thickness of 20 μm . Using parameters matched to the experiment (see Methods), we are able to reproduce the experimental results.

Figure 2 shows the energy spectra of the C^{5+} (green) and Pd^{21+} (red) components obtained in the simulation. The energy per nucleon of the C^{5+} ions agrees well with the measured energies and the total number of ions accelerated, albeit with a somewhat smaller energy spread. The energy spectrum and peak ionization state of Pd between 50 and 200 MeV are likewise in good agreement with the data. From this simulation we can also infer a source layer thickness of $\sim 10 \text{ Å}$ (that is, a few monolayers), which is in good agreement with published measurements—for example, using Auger spectroscopy^{24,25}. In our parametric simulation study, increasing the number of initial carbon layers while keeping the density fixed leads to two effects, shown in Fig. 3. The mean energy of the C^{5+} beam decreases and the energy spread increases with increased layer thickness. This trend continues until adequate space charge exists in the carbon layer to shield the ionizing electric field experienced by the carbon ions at the back of the layer. These more deeply buried carbon ions only attain ionization state C^{4+} and they separate from the layer of C^{5+} ions; for the parameters considered in this study, this occurs for areal charge densities exceeding $\sim 2 \times 10^{-8} \text{ g cm}^{-2}$, corresponding to a layer thickness of $\sim 80 \text{ Å}$ and above. Lower carbon ionization states appear with increasing layer thickness, and the ion energies eventually approach a maxwellian distribution. This behaviour is also seen in the experiment (Fig. 2). The grey curve shows the C^{4+} spectrum from a laser shot of comparable energy but from a tungsten target, which is not a catalyst for the required surface chemistry, and therefore does not form a thin source layer.

Measurements using transmission electron microscopy (see Supplementary Information section SI_4) reveal that upon heating, the target actually forms a 400-Å-thick tungsten carbide (W_2C) layer. This surface layer is not thin enough, and as a consequence the C spectrum is maxwellian and all lower charge states are present, as observed in earlier experiments⁶. Comparison of the two spectra shows that the direct production of monoenergetic ions by thin source layers is more effective than just slicing the equivalent energy range out of the maxwellian spectrum. Specifically, the number of ions in the corresponding energy range from 2.5 to 3.5 MeV per nucleon is a factor of 2 lower than in the monoenergetic case. Comparison with a cold Pd target shot at similar laser conditions also shows a maxwellian distribution of considerable lower energy (magenta curve, Fig. 2), because (1) the protons drain energy; and (2) the localized source layer is not formed. Our model predicts that the energy spread in the carbon beam may be minimized by localizing the initial carbon layer spatially, that is, by minimizing the source layer thickness, a process which should also result in higher mean energy of the light ion beam. This hypothesis will be tested in future experiments.

Our experimental results, simulation and analytic modelling have

established the basis for laser-driven acceleration of monoenergetic ion beams using specifically designed and treated targets. Moreover, catalytic metal substrates such as Pd offer the chance of having a target that configures itself *in situ* if subjected to the right conditions. Such a target would solve major technical obstacles for a host of possible applications, making future laser-based accelerators much more feasible. We recently confirmed these results by repeating the experiments in another campaign at the Trident facility, using a substantially equivalent experimental set-up, where we reproduced the qualitative findings reported here. Although the errors in the analysis for any specific shot are small, the reproducibility of our C^{5+} results from shot to shot is only $\sim 50\%$, possibly owing to the degree of control and diagnosis of key input parameters achievable in our present experimental set-up. Large, high-energy, single-shot glass lasers have typical shot-to-shot power fluctuations of $\sim 25\%$, and the focal spot conditions drift over time. Varying preplasma conditions and possible self-focusing add further to the variability of the results.

The resulting unique beam characteristics, including short pulse duration, high current and small transverse and longitudinal emittances, represent a strong incentive to pursue further research and applications, such as advanced accelerator concepts¹¹, laboratory astrophysics, isochoric heating²¹, fusion science¹⁵ and medical physics¹⁴. The achieved particle energy is already in the right energy range for fusion applications like fast ignition, whereas particle number and conversion efficiency have to be substantially increased. For medical applications like tumour therapy the situation is the opposite: here, the particle numbers are sufficient but the particle energy has to be increased substantially. Considering the fast paced progress in ultrahigh intensity laser technology in recent years, it is reasonable to anticipate progress on all these issues and the deployment of a laser-driven, quasi-monoenergetic ion accelerator in the not so distant future. Progress made in diode-pumped glass laser systems, especially, should enable far higher repetition rates of 0.1–1 Hz (ref. 26). At these repetition rates, several applications in accelerator physics, medical physics, material science and neutron physics become feasible.

METHODS

Laser system and diagnostics. The experiments were performed at the short pulse arm of the Trident Nd:glass laser facility at Los Alamos National Laboratory. The Trident C-beam delivers up to 30 TW in a 20 J, $\sim 600 \text{ fs}$ pulse at $1.054 \mu\text{m}$ wavelength, using chirped pulse amplification²⁷. The typical pulse contrast is $\sim 10^{-6}$ at 2 ns before the peak of the pulse. As illustrated in Fig. 1, an off-axis parabolic mirror is used to focus the laser pulse onto a thin foil target at 22.5° with respect to the target normal. Typical focal spot sizes are $\sim 10 \mu\text{m}$ radius, resulting in intensities on target of $\sim 10^{19} \text{ W cm}^{-2}$. A stack of Gafcom radiochromic film (RCF) is placed behind the target to record the ion beam profile. A hole in the middle of this film stack provides a line of sight for a Thomson parabola (TP) ion spectrometer²⁸ attached to the outer chamber wall. The Thomson parabola deflects the ions by means of parallel electric and magnetic fields, so that the projection of their path in the detector is defined by parabolic traces. Ions with different charge-to-mass ratios are deflected onto different traces, while their positions on a given trace are determined by their energies. A CR-39 solid state nuclear track detector records the ions, typically $\sim 300,000$ per shot, and is read out by a specialized automated analysis system²⁹. With properly chosen parameters, the counting error is below 0.01%. The error in ion numbers per energy bin (dN) is dominated by Poisson statistics, and is below $\leq 1\%$ owing to the large number of counts per shot. For example, for the C^{5+} trace in Fig. 2, $dN \approx 0.3\%$. The solid angles of the TPs are 3.4×10^{-5} milliradians (msr), and the opening angles of the ion beams are 24–100 msr depending on charge state and energy. The TPs are absolutely calibrated for energy and the energy error is dominated by the pinhole size (100 μm). It is given as $dE \sim E^{3/2}$, yielding an upper boundary for an energy error of less than 1 MeV for $\sim 45 \text{ MeV}$ carbon, that is, less than 2% and decreasing with energy.

Target treatment and chemistry. The foil target is heated to $T_f \approx 1,100 \text{ K}$ by two attached wires that pass a current through the foil. Palladium at room temperature is a hydrogen-getter, that is, H can be found throughout the bulk of the material as well as on the surfaces. The heating process desorbs the hydrogen contaminants (adsorbed and absorbed in the foil), thus enabling the efficient acceleration of heavier ions. In the experiment presented here, the special

catalytic surface chemistry of palladium causes a few carbon monolayers of hydrocarbon contaminants to remain on the surface of the palladium substrate and form a well defined source layer for the monoenergetic carbon beam. Given the ambient vacuum of $\sim 10^{-6}$ mbar, the surface is contaminated with various C_xH_y compounds. When the Pd is heated, the Pd surface undergoes multiple phase changes^{24,25} and the loosely bound H is driven out of the bulk and off the surfaces. At 600 K the target is completely dehydrogenized. The carbon, however, remains on the surface in various different configurations. When heating the target further, to temperatures $T > 1,100$ K, the various carbon compounds undergo a phase change, forming a well-defined, very thin graphite layer at the monolayer scale on the Pd surface. If heated up further, to above 1,300 K, this layer will be removed and a clean Pd surface remains. In the experiment, we did not reach this last state, but remained in the graphite regime, thereby preparing a thin source layer perfect for creating monoenergetic ions.

BILBO hybrid code. In BILBO, ion formation and acceleration is accomplished by the electric fields of a virtual cathode of hot electrons at the back surface of the target. Assuming separation of the electron and ion timescales, self-consistent electric fields are obtained by solving the time-stationary relativistic Vlasov–Maxwell equations for each electron component. These fields accelerate the ions and ionize them to higher charge states, where ionization is implemented in BILBO by means of a threshold ionization model³⁰. The boundary conditions require the electric field to vanish within the target and far from the target surface. In addition, the electron densities and temperatures of the hot and cold components are specified within the target as internal boundary conditions. The hot electron density and temperature are functions of the laser energy deposition model, and their dynamics include adiabatic expansion and the loss of energy to ionization and ion acceleration. The cold electron temperature increases from ohmic heating and collisions with the hot electron component. In the simulations, the laser spot diameter was assumed to be $30\text{ }\mu\text{m}$; 50% absorption of the incident laser into hot electrons was assumed ($T_h \approx 2.5\text{ MeV}$), with the hot electrons' density assumed to be equal to the critical density ($n_e = 1.01 \times 10^{21}$). The cold electrons had $n_c = 6.8 \times 10^{22}\text{ cm}^{-3}$ and initial cold electron temperature $T_c = 10\text{ eV}$. The density and temperature profiles of the hot electron component were assumed to evolve in time with gaussian shape during the pulse rise and have a full-width at half-maximum of 700 fs. The simulation used 5×10^4 simulation ions of each species, had a time step of 2 fs, and employed 6×10^5 simulation cells over a domain of size $100\text{ }\mu\text{m}$.

Received 17 August; accepted 3 November 2005.

- Katsouleas, T. Accelerator physics: Electrons hang ten on laser wake. *Nature* **431**, 515–516 (2004).
- Faure, J. *et al.* A laser–plasma accelerator producing monoenergetic electron beams. *Nature* **431**, 541–544 (2004).
- Mangles, S. *et al.* Monoenergetic beams of relativistic electrons from intense laser–plasma interactions. *Nature* **431**, 535–538 (2004).
- Geddes, C. *et al.* High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding. *Nature* **431**, 538–541 (2004).
- Hatchett, S. *et al.* Electron, photon, and ion beams from the relativistic interaction of Petawatt laser pulses with solid targets. *Phys. Plasmas* **5**, 2076–2082 (2000).
- Hegelich, M. *et al.* MeV ion jets from short-pulse-laser interaction with thin foils. *Phys. Rev. Lett.* **89**, 085002 (2002).
- Roth, M. *et al.* Energetic ions generated by laser pulses: A detailed study on target properties. *Phys. Rev. Spec. Topic Accelerators Beams* **5**, 061002 (2002).
- Snively, R. *et al.* Intense high-energy proton beams from petawatt-laser irradiation of solids. *Phys. Rev. Lett.* **85**, 2945–2948 (2000).
- Maksimchuk, A., Gu, S., Flippo, K., Umstadter, D. & Bychenkov, V. Yu. Forward ion acceleration in thin films driven by a high-intensity laser. *Phys. Rev. Lett.* **84**, 4108–4111 (2000).
- Cowan, T. *et al.* Ultralow emittance, multi-MeV proton beams from a laser virtual-cathode plasma accelerator. *Phys. Rev. Lett.* **92**, 204801 (2004).
- Habs, D., Pretzler, G., Pukhov, A. & Meyer-ter-Vehn, J. Laser acceleration of electrons and ions and intense secondary particle generation. *Prog. Part. Nucl. Phys.* **46**, 375–377 (2001).
- Cobble, J. A., Johnson, R. P., Cowan, T. E., Renard-Le Galloudec, N. & Allen, M. High resolution laser-driven proton radiography. *J. Appl. Phys.* **92**, 1775–1779 (2002).
- Borghesi, M. *et al.* Proton imaging: a diagnostic for inertial confinement fusion/fast ignitor studies. *Plasma Phys. Control Fusion* **43**, A267–A276 (2001).
- Ledingham, K. *et al.* High power laser production of short-lived isotopes for positron emission tomography. *J. Phys. D* **37**, 2341–2345 (2004).
- Kodama, R. *et al.* Fast heating of ultrahigh-density plasma as a step towards laser fusion ignition. *Nature* **412**, 798–802 (2001).
- Roth, M. *et al.* Fast ignition by intense laser-accelerated proton beams. *Phys. Rev. Lett.* **86**, 436–439 (2001).
- Temporal, M., Honrubia, J. J. & Atzeni, S. Numerical study of fast ignition of ablatively imploded deuterium–tritium fusion capsules by ultra-intense proton beams. *Phys. Plasmas* **9**, 3098–3107 (2002).
- Wilks, S. *et al.* Energetic proton generation in ultra-intense laser–solid interactions. *Phys. Plasmas* **8**, 542–549 (2001).
- Atzeni, S. & Meyer-ter-Vehn, J. *The Physics of Inertial Confinement Fusion* (Oxford Univ. Press, Oxford, 2004).
- Hegelich, M. *et al.* Spectral properties of laser-accelerated mid-Z MeV/u ion beams. *Phys. Plasmas* **12**, 056314 (2005).
- Patel, P. *et al.* Isochoric heating of solid-density matter with an ultrafast proton beam. *Phys. Rev. Lett.* **91**, 125004 (2003).
- Gitomer, S. J. *et al.* Fast ions and hot electrons in the laser–plasma interaction. *Phys. Fluids* **29**, 2679–2688 (1986).
- Esirkepov, T. Zh. *et al.* Proposed double-layer target for the generation of high-quality laser-accelerated ion beams. *Phys. Rev. Lett.* **89**, 175003 (2002).
- Hamilton, J. C. & Blakeley, J. M. Carbon segregation to single crystal surfaces of Pt, Pd and Co. *Surf. Sci.* **91**, 199–217 (1980).
- Ramsier, R. D., Lee, K.-W. & Yates, J. T. Jr A sensitive method for measuring adsorbed carbon on palladium surfaces: Titration by NO. *J. Vac. Sci. Technol.* **13**, 188–194 (1995).
- Hein, J. *et al.* Diode-pumped chirped pulse amplification to the joule level. *Appl. Phys. B* **79**, 419–422 (2004).
- Mourou, G. A. & Umstadter, D. Extreme light. *Sci. Am.* **286**, 80–86 (2002).
- Thomson, J. J. Rays of positive electricity. *Phil. Mag.* **21**, 225–249 (1911).
- Rusch, G., Winkel, E., Noll, A. & Heinrich, W. The Siegen automatic measuring system for track detectors: New developments. *Nucl. Tracks Radiat. Meas.* **19**, 261–265 (1991).
- Augst, S., Meyerhofer, D. D., Strickland, D. & Chint, S. L. Laser ionization of noble gases by Coulomb-barrier suppression. *J. Opt. Soc. Am. B* **8**, 858–867 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the expert support of the Trident laser team, especially R. Johnson, T. Ortiz and R. Gonzales, and the target fabrication support from LANL group MST-7, particularly R. Perea. This work was supported by the LANL Laboratory Directed Research & Development (LDRD) programme. One of the authors (H.R.) was supported by DOE/NNSA-UNR and another (J.S.) by DFG and BMBF.

Author Contributions B.M.H. conceived the experiment, B.M.H., J.C., S.L. and J.C.F. executed the experiment, B.M.H., J.S., K.F. and J.C.F. analysed the data, H.R., B.J.A. and B.M.H. did the theory, M.P. and R.K.S. helped with the material science part and palladium surface chemistry, and B.M.H., B.J.A. and J.C.F. wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.M.H. (hegelich@lanl.gov).

Laser-plasma acceleration of quasi-monoenergetic protons from microstructured targets

H. Schwöerer¹, S. Pfotenhauer¹, O. Jäckel¹, K.-U. Amthor¹, B. Liesfeld¹, W. Ziegler¹, R. Sauerbrey¹, K. W. D. Ledingham^{1,2,3} & T. Esirkepov^{4,5}

Particle acceleration based on high intensity laser systems (a process known as laser-plasma acceleration) has achieved high quality particle beams that compare favourably with conventional acceleration techniques in terms of emittance, brightness and pulse duration^{1–4}. A long-term difficulty associated with laser-plasma acceleration—the very broad, exponential energy spectrum of the emitted particles—has been overcome recently for electron beams^{5–7}. Here we report analogous results for ions, specifically the production of quasi-monoenergetic proton beams using laser-plasma accelerators. Reliable and reproducible laser-accelerated ion beams were achieved by intense laser irradiation of solid microstructured targets. This proof-of-principle experiment serves to illuminate the role of laser-generated plasmas as feasible particle sources. Scalability studies show that, owing to their compact size and reasonable cost, such table-top laser systems with high repetition rates could contribute to the development of new generations of particle injectors that may be suitable for medical proton therapy^{8–10}.

Beams of electrons, protons, ions and high energy photons produced by the interaction of ultraintense, ultrashort laser pulses with matter have received considerable attention throughout the last few years because of their high beam quality and their extensive potential for applications in science and technology^{8–10}.

The interaction of an intense light field with matter yields the generation of a hot plasma and the subsequent acceleration of electrons up to relativistic energies^{11,12}. Protons and ions are accelerated by a well controlled mechanism known as ‘target normal sheath acceleration’ (TNSA)¹³ following the initial electron acceleration (Fig. 1). Fast electrons are accelerated by an intense laser pulse (intensity $I \geq 10^{19} \text{ W cm}^{-2}$) from the surface of a thin metal foil in the forward direction. They penetrate the foil and ionize atoms along their paths. Within about a picosecond, those electrons leaving the target at the rear surface (that is, the back surface with respect to laser irradiation) build up a quasi-static electric field. The field acts normally to the target surface, has cylindrical symmetry and decreases in the transverse direction. Owing to the ultrashort duration of the electron bunch and its high charge, this field may reach values of several TV (10^{12} V m^{-1}) close to the axis and thus the potential can attain several tens of MeV (ref. 3). Protons and positively charged ions present on the back surface of the foil may be accelerated by this field until they compensate the electron charge. In most cases, the origin of these parasitic protons has been identified to be a hydrocarbon contamination layer on the target surface^{14,15}.

As the duration of the acceleration is ultrashort and the protons (as well as the ions) are at rest before acceleration, comprising a very small phase space volume, the transverse emittance of the proton beam reaches values as low as a few 10^{-3} mm mrad for 10 MeV

protons⁴. However to date, laser accelerated ion beams still show a large longitudinal emittance, with their energy spectrum exhibiting a quasi-exponential shape with a distinct cut-off energy^{1–3}. This can be explained by the inhomogeneous distribution of electrons in the sheath causing an accelerating field that is inhomogeneous in the transverse direction. For a plane and unstructured target, the transverse dimension of the electric field and hence the source size of accelerated protons are much larger than the laser’s focal spot¹⁶ (Fig. 1). Therefore, different parasitic protons experience a range of potentials, resulting in a broad distribution of energies.

Following this understanding of the mechanism of laser acceleration of protons, Bulanov *et al.*^{10,17} pointed out that the resulting proton energy spectrum has a strong correlation to the spatial distribution of the protons on the target surface. In order to generate high quality proton beams with monoenergetic features, they proposed a bilayered, microstructured target, consisting of a thin high Z

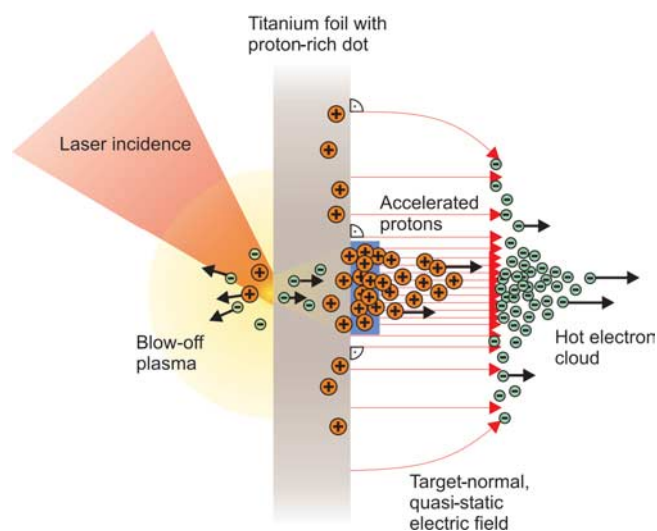


Figure 1 | Laser acceleration of protons from the back side of a microstructured target. A terawatt (TW)-laser pulse is focused onto the front side of the target foil, where it generates a blow-off plasma and subsequently accelerates electrons. The electrons penetrate the foil, ionize hydrogen and other atoms at the back surface and set up a Debye sheath. The inhomogeneous distribution of the hot electron cloud causes a transversely inhomogeneous accelerating field (target normal sheath acceleration—TNSA). Applying a small hydrogen-rich dot on the back surface enhances the proton yield in the central part of the accelerating field, where it is nearly homogenous. These protons constitute the quasi-monoenergetic bunch.

¹Institut für Optik und Quantenelektronik, Friedrich-Schiller-Universität, 07743 Jena, Germany. ²Department of Physics, University of Strathclyde, Glasgow G4 0NG, UK. ³AWE plc, Aldermaston, Reading RG7 4PR, UK. ⁴Kansai Research Establishment, JAERI, Kizu, Kyoto, 619-0215, Japan. ⁵Moscow Institute of Physics and Technology, Dolgoprudnyi, 141700, Russia.

metal foil and a small proton-rich dot on the back surface. The transverse dimension of such dots is smaller than that of the acceleration sheath, and hence the protons will only be subject to the central—that is, homogeneous—part of the acceleration field. In this configuration, the protons all experience the same electric field and are accelerated in the same potential (Fig. 1). The resulting proton beam has a spectrum with a strong monoenergetic peak.

Here we present experimental results of a laser accelerated proton beam with a peaked energy distribution. The experimental arrangement follows the proposal in ref. 17 to use microstructured targets (Fig. 2). The results are well reproduced by two-dimensional particle-in-cell (PIC) simulations. The simulations also show the scalability of the technique.

The high intensity laser pulses (intensity $I = 3 \times 10^{19} \text{ W cm}^{-2}$) are generated by the JETI 10 TW Ti:sapphire laser at the University of Jena. It delivers pulses of 80 fs duration, a pulse energy of 600 mJ on target and a maximum repetition rate of 10 Hz. The target is a 5 μm thin titanium foil, coated with a 0.5 μm layer of polymethyl methacrylate (PMMA) on the back surface. In some regions of the sample the PMMA layer was microstructured, leaving PMMA dots of $(20 \times 20) \mu\text{m}^2$ on the surface with PMMA-free space around (Fig. 2b). The laser pulse hits the foil on the front surface exactly opposite to one of the dots. Protons and ions, accelerated from the back surface in the normal direction, were analysed by a Thomson spectrometer and detected either by an online system

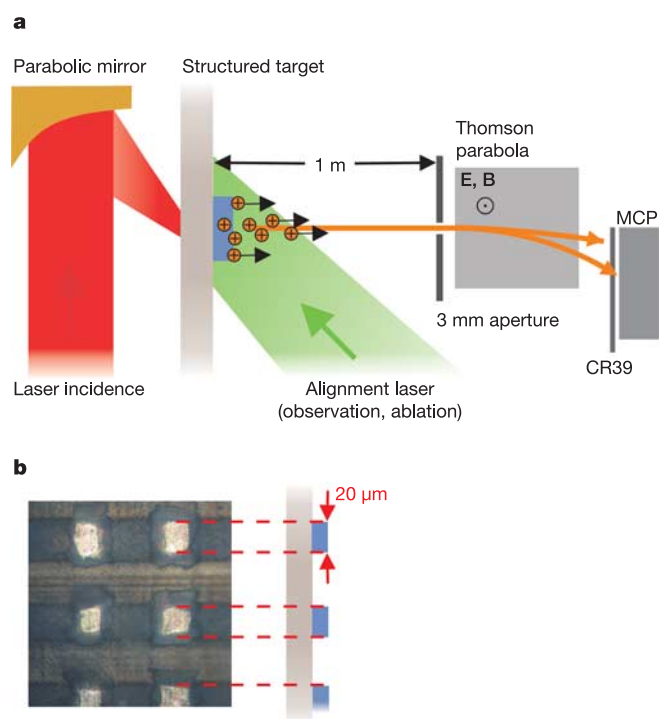


Figure 2 | Experimental and target arrangement for laser proton acceleration from microstructured targets. **a**, Experimental set-up: a TW-laser pulse is focused by a 45° off-axis parabolic mirror ($f/2.5$) to an intensity of $3 \times 10^{19} \text{ W cm}^{-2}$. A dot on the back surface of the target foil is positioned opposite to the focus using an alignment laser. A titanium foil of 5 μm thickness carries dots of PMMA with a thickness of 0.5 μm and a transverse size of $20 \mu\text{m} \times 20 \mu\text{m}$. The photograph of the microstructured back surface in **b** shows the dots as light squares. The protons and ions accelerated from the target are dispersed with respect to energy and charge-to-mass ratio in a Thomson parabola, and then detected by either a microchannel plate (MCP) with phosphor screen and CCD camera or on nuclear track detector plastics (CR39). A 3 mm lead aperture in front of the Thomson parabola serves as a pinhole for the Thomson spectrometer and additionally shields the MCP from bremsstrahlung.

based on microchannel plates (MCP) or by nuclear track detectors (CR39).

Figure 3 presents the result of irradiating the microstructured target foil at the position of a dot in contrast to unstructured material. The data are given as number of protons per energy interval of 0.05 MeV, which corresponds to our MCP resolution, and per 24 msr, which is the solid angle of emission, that results from the simulations described below. The curve indicated by the blue triangles shows the proton spectrum after irradiating a dot. It exhibits a distinct narrow band feature, peaked around $E_{\text{max}} = 1.2$ MeV on top of a broad, exponential shaped background. The displayed feature contains about 10^8 protons per 24 msr, and has a full-width at half-maximum of about $\Delta E_{\text{FWHM}} = 300 \text{ keV}$ or 25% of its absolute value. For comparison, the black data represent an average over six proton spectra recorded if the laser hits a blank position on the same target, where protons can only originate from an unstructured hydrogen contamination layer (parasitic protons). No narrow band feature appears, and the exponential shape of the spectrum can be approximated by a temperature of about 0.5 MeV. The occurrence of the peaked spectrum after irradiating a dot was also observed using CR39 detection. In the inset in Fig. 3, spectra obtained with the two detection methods are compared under the same conditions. Red circles and black squares represent spectra using MCP and CR39 detection, respectively. The two curves exhibit the same shape with a peak at 1.2 MeV, which accounts both for the reproducibility of generating peaked spectra as well as the reliability of the MCP detection. The proton yield at spectrum maximum is determined not only by the number of abundant protons in the dot but also by the laser characteristics, which affect the transverse scale of the electrostatic potential, the acceleration time and the spatial distribution of protons.

Owing to the limited size of the multichannel plate, the spectral range of the MCP data is smaller than that of the CR39 data sets. However, following direct successful comparison of these measurements, the electronic system provides a reliable online observation which allows for systematic investigation of the reproducibility with respect to proton yield, peak position and spectral width as well as for fast and controlled changes of experimental conditions. This

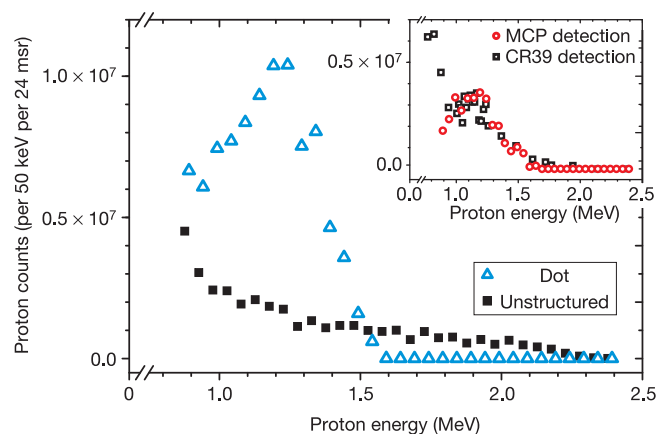


Figure 3 | Proton spectra from the Thomson spectrometer. The proton number reaching the detector is given per energy interval of 0.05 MeV and per solid angle of 24 msr versus proton energy. The main graph shows a spectrum obtained from irradiating the foil at the position of a dot using the MCP detection. It is represented by blue triangles. The spectrum from a dot exhibits a peak at an energy of 1.2 MeV as opposed to exponential spectra (black squares, average from six shots) in the case of using an unstructured part of the target foil. The peaked structure contains about 10^8 protons per 24 msr. The inset shows the comparison of the two detection systems: both spectra show the energy distribution of the protons from irradiating a dot under the same conditions. Red circles represent MCP detection, while black squares originate from CR39 detection.

experimental accuracy can only be accomplished with table-top lasers with high repetition rate. A maximum in the proton spectrum is reproduced consistently if a microstructure on the rear of the target is irradiated. The position of the peak E_{max} as well as its width ΔE_{FWHM} , vary from shot to shot by $\pm 20\%$.

From systematic experiments performed earlier on non-microstructured targets, it is known that an optimum target thickness with respect to maximum proton energy and yield follows from a given temporal structure of the laser pulse¹⁸. Considering our laser and target conditions, the exponential contribution to the proton distribution from the non-microstructured targets is in agreement with previously published results from similar targets^{18–20}.

The narrow band spectra of the laser accelerated protons observed in our experiments are due to the small proton-rich area within the centre of the larger quasi-static electric field, set up by the laser accelerated electrons beyond the thin target (Fig. 1). If the scale of the inhomogeneity of the electric field is larger than the proton-rich spot, these protons all experience the same potential. The maximum proton energy is determined by the total charge of electrons constituting the acceleration sheath and occurs on the axis, which in turn depends on the laser intensity and the target thickness¹. This analysis is supported by two-dimensional PIC multi-parametric simulations based on the code REMP and the model of ref. 20: Fig. 4 plots the resulting proton spectrum (solid line) under the conditions laser intensity $I = 3 \times 10^{19} \text{ W cm}^{-2}$ and a $5 \mu\text{m}$ thin titanium target with $20 \mu\text{m}$ PMMA structure on the back side, which match the experimental parameters. The numerically achieved proton spectrum is dominated by a narrow band structure around 1.2 MeV with a full-width at half-maximum $\text{FWHM} = 0.3 \text{ MeV}$ or 25%. Simulation and experiment are in good agreement with respect to both the existence of the narrow structure as well as its position and width.

It was estimated^{10,17} that in the case of a $10 \mu\text{m}$ focal spot and $0.1 \mu\text{m}$ thick proton layer, a proton energy spread of 1% can be expected. Furthermore, for microstructured targets the maximum proton energy (E_{max} in MeV) scales as the square root of the laser power (P in PW) with $E_{\text{max}} = 230 \times P^{1/2}$ (refs 21, 22). The

parameters for the simulation have been changed to smaller dot sizes and higher energies in order to extrapolate the technique to future experiments and investigate the scalability of our results (see Fig. 4 inset). A high repetition rate table-top laser system with petawatt power (POLARIS) will be available within a few years²³. This laser will deliver pulses of 150 J within 150 fs , which leads to an intensity of about $10^{21} \text{ W cm}^{-2}$ in focus. Simulations performed with these parameters result in a peak proton energy at 173 MeV and relative width $\Delta E/E \approx 1\%$ for a dot diameter of $2.5 \mu\text{m}$ and a reduced layer thickness of $0.1 \mu\text{m}$. Under these conditions, the proton yield is not longer limited by laser energy, but all protons contained in the dot (8×10^8) are quickly accelerated to an energy of $173 \pm 1 \text{ MeV}$.

Proton beams with such a characteristic might be suitable for treatment of deep seated tumours¹⁰. More realistic within a few years is the potential for laser induced proton therapy for eye tumours, which only requires $60\text{--}70 \text{ MeV}$ protons²⁴.

We have indicated that this experiment was carried out as ‘proof of principle’. An extensive programme is now underway to reduce the width and increase the energy of our proton peaks by improving the target fabrication procedure. We intend to reduce the dimensions of the dots and change the target material. Our initial targets were dots on titanium, which was chosen on the basis of our preliminary experiments. Gold is expected to be a much better substrate¹⁷, as it can deliver more electrons. In the Methods section, we describe how a Nd:YAG laser was used to reduce the thickness of the parasitic (contamination) layers by laser ablation²⁵. Also, the same laser illuminated the dots to align their position with respect to the focal region. This straightforward optical positioning will allow future operation of the set-up at high repetition rate.

We have demonstrated the feasibility of laser–plasma accelerators for producing proton spectra with a reproducible monoenergetic peak using microstructured targets. Because of the high spectral density, all accelerated protons can be allocated to the respective application with high conversion efficiency. Our first steps towards monoenergetic protons show a distinct improvement over the exponential energy spectra published to date, heralding new possibilities for ion injectors and compact accelerators. In the longer term, future laser accelerators may be in reach of medical proton and heavy ion therapy.

METHODS

Laser. The experiments were performed using the JETI laser, a multi-TW Ti:sapphire laser system based on chirped pulse amplification. It delivered laser pulses containing $600 \pm 25 \text{ mJ}$ on target within a pulse duration of $80 \pm 5 \text{ fs}$ ($1 \text{ fs} = 10^{-15} \text{ s}$). The pulse duration was determined using a background-free THG autocorrelator and a Mach–Zehnder pulse front interferometer. Additionally, the amplified spontaneous emission was suppressed to a contrast of $10^8/1$ at a time of 0.5 ps before the arrival of the main laser pulse using a fast Pockels cell. JETI operates at a centre wavelength of 795 nm and delivers a maximum of 10 shots per second. The laser parameters were fully characterized and monitored during the experiment. The laser pulses were focused by a 45° off-axis parabolic mirror ($f/2.5$) to a focal spot size of $7 \mu\text{m}^2$ (radius $r = 1.5 \mu\text{m}$), yielding an intensity of $3 \times 10^{19} \text{ W cm}^{-2}$.

Target. In the experiments presented here, the solid state target was a $5 \mu\text{m}$ thin titanium foil, which was coated on one side (the back side with respect to laser irradiation) with a $0.5 \mu\text{m}$ thin layer of PMMA doped with Rhodamine 6G. The suitable thicknesses of the metal foil ($5 \mu\text{m}$) as well as the PMMA layer ($0.5 \mu\text{m}$) with respect to proton energy and proton number was determined in preliminary experiments. The $5 \mu\text{m}$ titanium foil was chosen for reasons of stability and reproducibility, even though thinner metal foils ($2 \mu\text{m}$) resulted in higher proton energies. The hydrogen-rich PMMA layer was microstructured by means of femtosecond laser ablation, yielding dots with a size of $(20 \times 20) \mu\text{m}^2$ with a ‘PMMA-free’ space of $50 \mu\text{m}$ in between, see Fig. 2b. The target was stretched in a rigid frame to ensure optimal flatness and moved before each shot to an unirradiated, that is, intact, area. In the key part of the experiments each laser pulse was launched on the front side of the target exactly opposite to one of the dots. This was accomplished by localizing the dots with help of a frequency doubled Nd:YAG laser (532 nm) and observing the yellow fluorescence light (around 600 nm) of the dot. The foci of the TW pulse and the Nd:YAG laser were aligned to overlap in space. The dot’s position, determined from the back side,

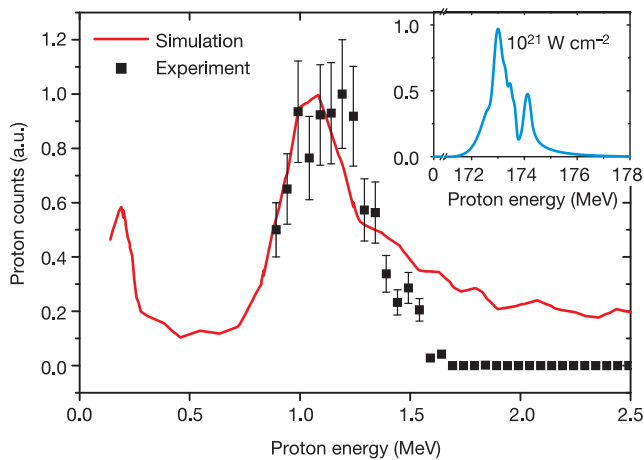


Figure 4 | Results from simulations and scalability of the technique. Comparison of experimental data (black squares) to the proton spectrum obtained from two-dimensional-PIC simulation (red line) for following conditions: laser intensity $I_L = 3 \times 10^{19} \text{ W cm}^{-2}$, and target dimensions $5 \mu\text{m}$ Ti foil + $0.5 \mu\text{m}$ PMMA dot ($20 \times 20 \mu\text{m}^2$). Experimental data points comprise the observable energy range on the MCP detector. The statistical uncertainty for the measured data has a value of 20% s.d. as shown by the error bars. In the inset a simulation for a petawatt-laser system demonstrating the scalability of proton acceleration from microstructured targets is shown. The parameters for the simulation are $I_L = 1.2 \times 10^{21} \text{ W cm}^{-2}$, $5 \mu\text{m}$ Ti foil + $0.1 \mu\text{m}$ PMMA dot ($2.5 \mu\text{m}$ diameter). The proton spectrum exhibits a narrow peak with relative energy width of $\Delta E/E \approx 1\%$ at a peak energy of 173 MeV .

was brought into spatial coincidence with the position of the focal spot on the front side. An additional effect of the irradiation of the target's back surface with the Nd:YAG laser was a reduction of the parasitic proton layer by laser ablation before the high intensity laser struck the target.

Proton detection. Protons and ions accelerated from the foil were detected either by a chevron microchannel plate (MCP) with phosphor screen and CCD camera or on nuclear track detector CR39. Before hitting the detector, the proton/ion beam traversed a Thomson parabola type spectrometer comprising a magnetic field ($B = 535$ mT) parallel to an electric field ($E_{\text{max}} = 2,500$ V cm⁻¹). Thus, the proton/ion beam was dispersed with respect to energy and charge-to-mass ratio. The energy distribution was then determined from the spatial distribution of the protons on the detector. The MCP system has the major advantage of providing real time detection. It was carefully calibrated by a comparison of pixel counts to proton pits on a piece of CR39. The energy spectra from MCP and CR39 are measured with an energy resolution of 50 keV, which corresponds to the resolution of the MCP, and per solid angle of 10 μsr defined by the 3 mm aperture in front of the Thomson parabola. The spectra are scaled to 24 msr, which is the full angle of emission as obtained from the simulation.

Received 14 September; accepted 1 December 2005.

- Maksimchuk, A., Gu, S., Flippo, K., Umstadter, D. & Bychenkov, V. Forward ion acceleration in thin films driven by a high-intensity laser. *Phys. Rev. Lett.* **84**, 4108–4111 (2000).
- Clark, E. *et al.* Measurement of energetic proton transport through magnetized plasma from intense laser interactions with solids. *Phys. Rev. Lett.* **85**, 1654–1657 (2000).
- Snively, R. A. *et al.* Intense high-energy proton beams from petawatt-laser irradiation of solids. *Phys. Rev. Lett.* **85**, 2945–2948 (2000).
- Cowan, T. E. *et al.* Ultralow emittance, multi-MeV proton beams from a laser virtual-cathode plasma accelerator. *Phys. Rev. Lett.* **92**, 204801 (2004).
- Mangles, S. *et al.* Monoenergetic beams of relativistic electrons from intense laser-plasma interactions. *Nature* **431**, 535–538 (2004).
- Geddes, C. *et al.* High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding. *Nature* **431**, 538–541 (2004).
- Faure, J. *et al.* A laser-plasma accelerator producing monoenergetic electron beams. *Nature* **431**, 541–544 (2004).
- Tajima, T. & Mourou, G. Zetawatt-exawatt lasers and their applications in ultrastrong-field physics. *Phys. Rev. Spec. Topics Accelerators Beams* **5**, 031301 (2002).
- Ledingham, K. W. D., McKenna, P. & Singhal, R. P. Applications for nuclear phenomena generated by ultra-intense lasers. *Science* **300**, 1107–1111 (2003).
- Bulanov, S. & Khoroshkov, V. Feasibility of using laser ion accelerators in proton therapy. *Plasma Phys. Rep.* **28**, 453–456 (2002).
- Modena, A. *et al.* Electron acceleration from the breaking of relativistic plasma waves. *Nature* **377**, 606–608 (1995).
- Malka, V. *et al.* Electron acceleration by a wake field forced by an intense ultrashort laser pulse. *Science* **298**, 1596–1600 (2002).
- Wilks, S. *et al.* Energetic proton generation in ultra-intense laser-solid interactions. *Phys. Plasmas* **8**, 542–549 (2001).
- Gitomer, S. J. *et al.* Fast ions and hot-electrons in the laser-plasma interaction. *Phys. Fluids* **29**, 2679–2688 (1986).
- Hegelich, M. *et al.* MeV ion jets from short-pulse-laser interaction with thin foils. *Phys. Rev. Lett.* **89**, 085002 (2002).
- Schreiber, J. *et al.* Source-size measurements and charge distributions of ions accelerated from thin foils irradiated by high-intensity laser pulses. *Appl. Phys. B* **79**, 1041–1045 (2004).
- Esirkepov, T. *et al.* Proposed double-layer target for the generation of high quality laser accelerated ion beams. *Phys. Rev. Lett.* **89**, 175003 (2002).
- Kaluza, M. *et al.* Influence of the laser prepulse on proton acceleration in thin-foil experiments. *Phys. Rev. Lett.* **93**, 045003 (2004).
- Spencer, I. *et al.* Experimental study of proton emission from 60-fs, 200-mJ high-repetition-rate tabletop-laser pulses interacting with solid targets. *Phys. Rev. E* **67**, 046402 (2003).
- Matsukado, K. *et al.* Energetic protons from a few-micron metallic foil evaporated by an intense laser pulse. *Phys. Rev. Lett.* **91**, 215001 (2003).
- Bulanov, S. V. *et al.* Feasibility of using laser acceleration in proton therapy. *AIP Conf. Proc.* **740**, 414–429 (2004).
- Esirkepov, T. *et al.* Laser ion acceleration scaling laws seen in multi-parametric PIC simulations. *Phys. Rev. Lett.* (submitted); preprint at (<http://arxiv.org/abs/physics/0510189>) (2005).
- Hein, J. *et al.* Diode pumped chirped pulse amplification to the joule level. *Appl. Phys. B* **79**, 419–422 (2004).
- Damato, B., Kacperek, A., Chopra, M., Campbell, I. R. & Errington, R. D. Proton beam radiotherapy of choroidal melanoma: The Liverpool-Clatterbridge experience. *Int. J. Radiat. Oncol. Biol. Phys.* **62**, 1405–1411 (2005).
- Allen, M. *et al.* Direct experimental evidence of back-surface ion acceleration from laser-irradiated gold foils. *Phys. Rev. Lett.* **93**, 265004 (2004).

Acknowledgements This work was supported by the Deutsche Forschungsgemeinschaft. T.E. thanks S. V. Bulanov for discussions. K.W.D.L. takes pleasure in the receipt of a Carl-Zeiss visiting professorship. We thank F. Ronneberger and B. Beleites for their technical support. We thank H.-J. Fuchs and W. Gräf for their help in producing the targets.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.S. (schwoerer@iqo.uni-jena.de).

Amplification of chirality in two-dimensional enantiomorphous lattices

Roman Fasel¹, Manfred Parschau¹ & Karl-Heinz Ernst¹

The concept of chirality dates back to 1848, when Pasteur manually separated left-handed from right-handed sodium ammonium tartrate crystals¹. Crystallization is still an important means for separating chiral molecules into their two different mirror-image isomers (enantiomers)², yet remains poorly understood³. For example, there are no firm rules to predict whether a particular pair of chiral partners will follow the behaviour of the vast majority of chiral molecules and crystallize together as racemic crystals⁴, or as separate enantiomers. A somewhat simpler and more tractable version of this phenomenon is crystallization in two dimensions, such as the formation of surface structures by adsorbed molecules. The relatively simple spatial molecular arrangement of these systems makes it easier to study the effects of specific chiral interactions⁵; moreover, chiral assembly and recognition processes can be observed directly and with molecular resolution using scanning tunnelling microscopy^{6–9}. The enantio-separation of chiral molecules in two dimensions is expected to occur more readily because planar confinement excludes some bulk crystal symmetry elements and enhances chiral interactions^{10,11}; however, many surface structures have been found to be racemic^{12–18}. Here we show that the chiral hydrocarbon heptahelicene on a Cu(111) surface does not undergo two-dimensional spontaneous resolution into enantiomers¹⁹, but still shows enantiomorphism on a mesoscopic length scale that is readily amplified. That is, we observe formation of racemic heptahelicene domains with non-superimposable mirror-like lattice structures, with a small excess of one of the heptahelicene enantiomers suppressing the formation of one domain type. Similar to the induction of homochirality in achiral enantiomorphous monolayers²⁰ by a chiral modifier, a small enantiomeric excess suffices to ensure that the entire molecular monolayer consists of domains having only one of two possible, non-superimposable, mirror-like lattice structures.

Mixtures of M- and P-enantiomers of the helically shaped heptahelicene ($C_{30}H_{18}$, [7]H, Fig. 1a) have been deposited onto the (111)-surface of a copper single crystal (Supplementary Methods). The scanning tunnelling microscope (STM) image for a closed-packed molecular layer of racemic [7]H on Cu(111) in Fig. 1b shows extended two-dimensional (2D) domains with a diameter limited mainly by the terrace width of the underlying substrate surface (~ 50 nm). Enantiomorphism is manifested by the oblique alignment of the adsorbate lattice, which destroys the reflection symmetry of the underlying substrate. The lattice vectors of the two mirror-domains λ and ρ are tilted by opposite angles of $\pm 10.9^\circ$ with respect to the $[1\bar{1}0]$ substrate surface direction. This 2D enantiomorphism has previously been observed by means of low energy electron diffraction (LEED)¹⁹ and attributed to spontaneous resolution of the mixture into homochiral domains. However, the present STM observations reveal that the local molecular arrangement within the two domains

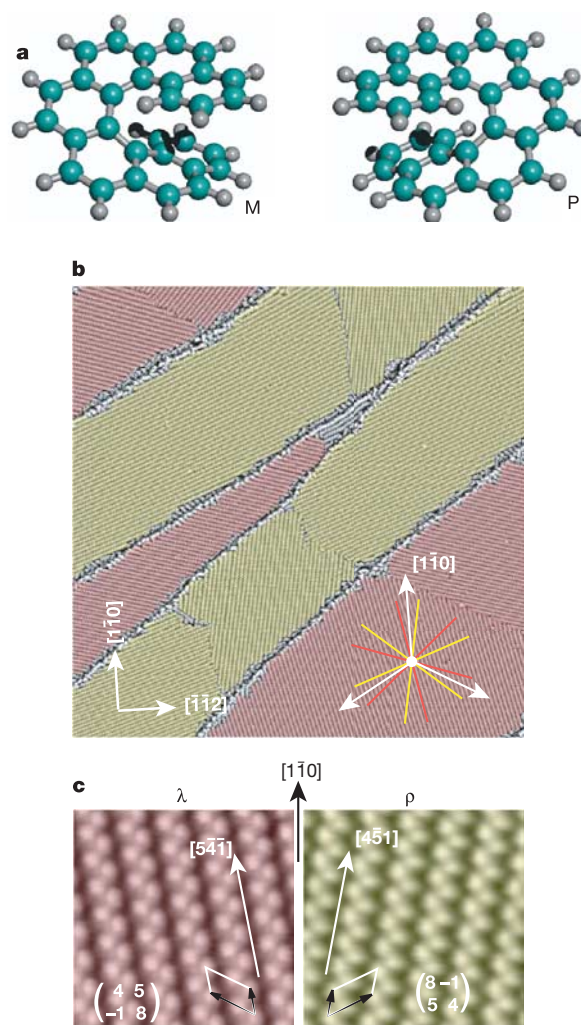


Figure 1 | Enantiomorphous domains of racemic heptahelicene on Cu(111). **a**, Ball and stick model of the two enantiomers of heptahelicene (green, carbon; grey, hydrogen). **b**, STM image ($200 \times 200 \text{ nm}^2$) showing the formation of extended close-packed domains. Left- (λ) and right-handed (ρ) domains are coloured red and yellow, respectively. **c**, Close-up images ($10 \times 10 \text{ nm}^2$) of the two mirror domains reveal zigzag rows running along the $[5\bar{4}1]$ (λ) and $[451]$ (ρ) directions (ρ), forming angles of $\pm 10.9^\circ$ with respect to the close-packed $[1\bar{1}0]$ surface directions. The $\begin{pmatrix} 4 & 5 \\ 8 & -1 \end{pmatrix}$ (λ domain) and $\begin{pmatrix} 8 & -1 \\ 5 & 4 \end{pmatrix}$ (ρ domain) unit cells are indicated.

¹Empa, Swiss Federal Laboratories for Materials Testing and Research, Überlandstrasse 129, CH-8600 Dübendorf, Switzerland.

(Fig. 1c) is strikingly different from the “pinwheel” and “cloverleaf” structures observed for the pure [7]H enantiomers on Cu(111)²¹, strongly indicating that spontaneous resolution has not occurred. Indeed, high-resolution STM images and molecular modelling show the domains to be racemic, with a local structure of regular zigzag rows of M- and P-enantiomers (Fig. 2).

But even though the formation of monolayers of racemic [7]H adsorbed on Cu(111) does not result in spontaneous resolution into 2D conglomerates, mirror-domains are nevertheless observed. It is known that self-assembly of chiral and achiral molecules at surfaces often leads to the formation of chiral motifs²². In the present case the heterochiral M-P pairs—although consisting of opposite enantiomers—represent truly chiral units (Fig. 2a,b) that break the symmetry of the system. The enantiomorphism in the lattice is thus not based only on the oblique orientation of the adsorbate mesh with respect to the substrate geometry: the molecular lattice of each racemic mirror-domain is itself homochiral. But because formation of left- (λ) and right-handed (ρ) domains is equally probable (Fig. 1), adsorption of the racemate results in an overall achiral surface.

This situation changes dramatically after deposition of [7]H with an enantiomeric excess (e.e. = $\{[M] - [P]\}/\{[M] + [P]\}$), that is, one enantiomer prevails on the surface (Fig. 3). At e.e. = 0.08, a ratio of 54% M-[7]H to 46% P-[7]H, exclusively λ domains are already observed. As discussed above, with respect to [7]H composition the

domains are racemic; that is, the λ domains formed contain both enantiomers at equimolar ratio. The excess M-[7]H must therefore accumulate outside the domains. With increasing enantiomeric excess, the average size of the perfectly ordered racemic domains decreases and residual areas containing small clusters and short zigzag rows increase accordingly (Fig. 3b). A fit of this linear increase confirms that the enantiomeric excess is not incorporated into the racemic λ domains but expelled to defects such as step edges and into the poorly ordered residual areas. This conclusion is further supported by the observation that for e.e. > 0.2, structural features characteristic for the pure enantiomers²¹ appear in residual areas close to step-edges (Supplementary Fig. 1). In addition, molecular mechanics calculations show that for steric reasons, the λ and ρ domains cannot incorporate an excess of either enantiomer of [7]H (Supplementary Fig. 2).

A first hint towards the likely origin of the highly nonlinear response of surface chirality to the enantiomeric excess of the deposited [7]H comes from the observation that boundaries between λ and ρ domains are rarely observed on flat terraces. In Figs 1b and 3

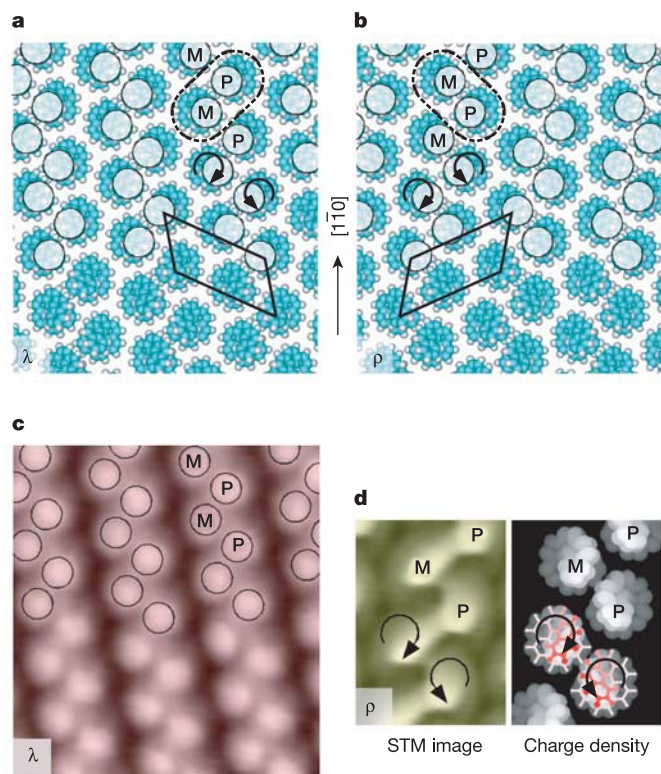


Figure 2 | Identification of enantiomeric composition. **a, b**, Lattice structures of the two lowest-energy configurations of racemic [7]H on Cu(111) obtained from molecular mechanics calculations. Pale circles highlight the topmost parts of the molecules expected to be imaged brightest via STM. The dashed ovals indicate heterochiral M-P pairs (molecule-molecule distance, 9.2 Å) forming the basic building blocks of the zigzag rows. **c**, STM image ($6.7 \times 6.2 \text{ nm}^2$) of a λ domain with the predicted locations of bright features from **a** superimposed. The molecular mechanics calculation perfectly matches the experimental STM image. **d**, High-resolution STM image ($3 \times 2.2 \text{ nm}^2$; gap voltage $U_{\text{gap}} = 1.6 \text{ V}$, tunnelling current $I_{\text{tunnel}} = 1 \text{ nA}$) of six [7]H molecules within a right-handed ρ domain (left) and extended-Hückel calculation of the LUMO to LUMO + 3 charge density (right). M- and P-configurations are clearly identified.

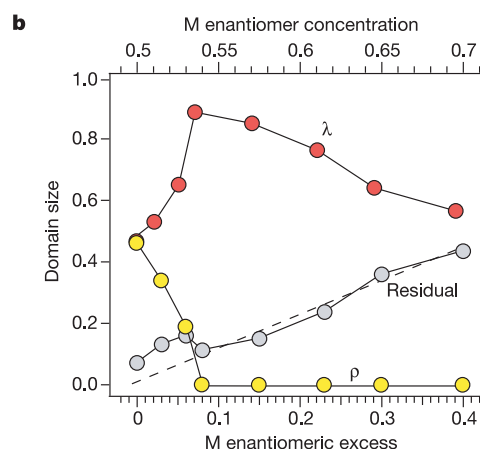
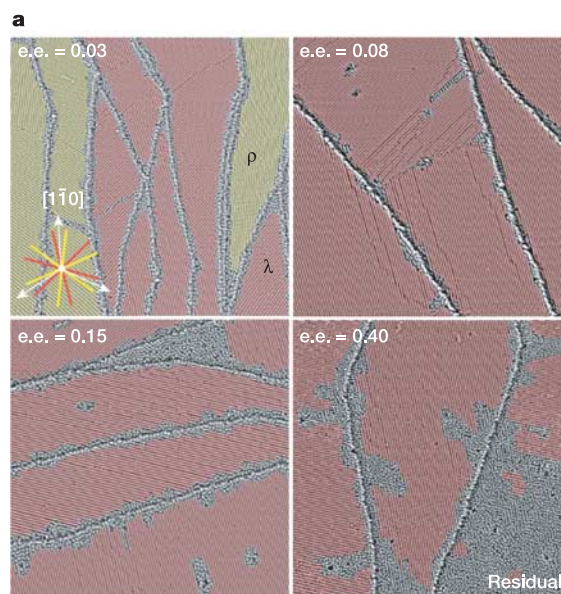


Figure 3 | Nonlinear amplification of chirality in non-racemic layers. **a**, Representative series of STM images ($200 \times 200 \text{ nm}^2$) for increasing enantiomeric excess. At e.e. = 0.08, the ρ domains (yellow) have completely disappeared and only λ domains (red) are observed. **b**, Normalized areas of λ (red circles) and ρ domains (yellow circles) for increasing enantiomeric excess, determined from a series of images such as those shown in **a**. Residual areas (grey circles) consisting of small clusters and short zigzag rows ('residuals') increase linearly with enantiomeric excess.

(e.e. = 0.03), a given substrate surface terrace is always occupied exclusively by a single domain type; boundaries between opposite mirror domains coincide with discontinuities such as steps edges. This suggests that boundaries between mirror domains are energetically sufficiently unfavourable that only a single mirror-domain will form on each terrace. Without a chiral bias, terraces will be covered either by λ or ρ domains (Fig. 1b). But if a small chiral bias favours the presence of one domain, the energy cost of mirror-domain boundaries might be sufficient to avoid formation of the opposite enantiomorphous lattice.

Large cooperative responses to a small chiral bias have been observed previously for helical co-polymers²³, where a small excess of chiral over achiral side chains drives the entire main polymer chain to adopt a single helicity. In analogy to the unfavourable energy cost of having a boundary between mirror domains, a helix reversal within the polymer chain imposes an unfavourable situation; the effect is amplified for further close-packing of the chains into parallel bundles²³. In our 2D system, the chiral bias arises from the enantiomeric excess located at the domain edges (Fig. 4): the oblique orientation of the λ and ρ domain lattices results in an energy difference between placing either M- or P-enantiomers at the domain edges. In other words, the interfaces between λ or ρ domains and adjacent M- or P-enantiomers are energetically non-equivalent. We observe that M-[7]H-excess favours λ domains, and P-[7]H excess favours ρ domains, and thus conclude that the corresponding M- λ and P- ρ interfaces are energetically preferred over M- ρ and P- λ interfaces.

The two energy factors driving the cooperative chiral order are thus (Fig. 4) the energy E^{MDB} that is associated with creating a boundary between the mirror domains λ and ρ (the mirror domain boundary energy), and the energy E^{DRI} that is associated with creating an interface between a mirror domain and a residual, enantio-enriched area (the domain/residual area interface energy). The interface energy $E^{\text{DRI}} = f(\text{e.e.})$ acts as a chiral field, which (depending on the handedness of the excess enantiomer) slightly favours one of the two possible domains over the other so that the

racemic lattice adjacent to the interface crystallizes into the low-energy orientation. This favoured domain with the low-energy orientation, in turn, will cover as large a surface area as possible in order to avoid energetically unfavourable mirror domain boundaries. Molecular mechanics calculations show that boundaries between mirror domains cost twice the energy of the frequently observed boundaries between rotational domains, that is, domains of the same handedness, but having their orientation with respect to the substrate rotated relative to each other (Supplementary Fig. 3). Overall, the E^{MDB} penalty thus ensures that many M-P dimers end up adopting an identical mirror domain orientation. In this way, the small chiral bias of E^{DRI} at the interface of the domains is amplified so that it drives the entire system into homochirality. Molecular mechanics calculations confirm that the chiral bias is indeed induced by excess molecules located at the domain edges: In all calculations, the λ domain edges interacting with excess M-enantiomers are energetically favoured over λ domain edges interacting with the opposite enantiomer (Supplementary Fig. 4), whereas ρ domains favour P-enantiomers at their edges. This effect is strongest at monolayer saturation coverage, when M-[7]H excess molecules next to a λ domain still interact attractively, while P-[7]H molecules impose a strongly repulsive situation.

Heptahelicene on Cu(111) is the first example of a system where a small enantiomeric excess induces homochirality. Although an identical effect can obviously not be observed in systems where adsorbed molecules separate into enantiopure domains, we expect the underlying boundary-driven amplification mechanism to be fundamental and relevant to a wide range of 2D systems. Homochiral templates created by exploiting boundary-driven chiral amplification might find use as asymmetric nucleation sites for further transfer of chirality into the third dimension, that is, into chiral macroscopic crystals. When combined with systems that allow triggering or switching of chiral bias in a surface monolayer (for example by optical methods), the boundary-driven chiral amplification effect we have described here might make it possible to dynamically control chirality in synthetic systems of possible interest for asymmetric chemical synthesis, chiral separation or the fabrication of advanced functional materials.

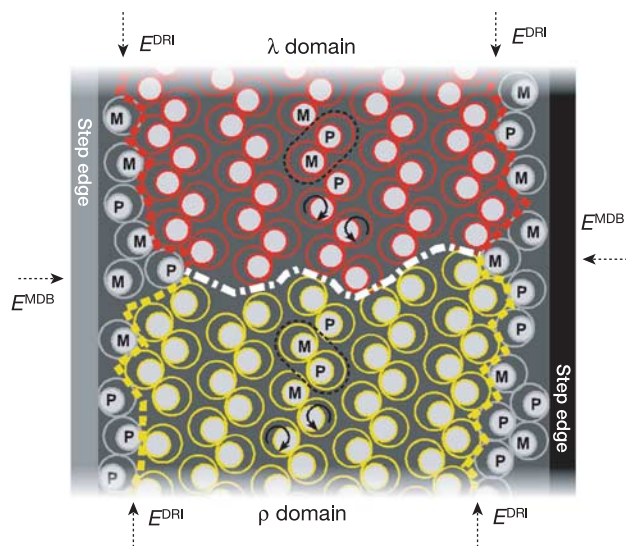


Figure 4 | Illustration of the energetic contributions driving the chiral amplification mechanism. Domain boundaries between λ and ρ domains (white dash-dotted line, E^{MDB}) are energetically unfavourable and rarely observed. A chiral bias, favouring one domain orientation, arises from the enantiomeric excess located at the interface between the perfectly ordered racemic domains and residual areas, at first developed along step edges (red and yellow dashed lines, domain/residual interface energy E^{DRI}). The sign of enantiomeric excess, that is, the handedness of the excess enantiomer, determines the favourable domain orientation.

Received 3 August; accepted 9 November 2005.

- Pasteur, L. Recherches sur les relations qui peuvent exister entre la forme cristalline et la composition chimique, et le sens de la polarisation rotatoire. *Ann. Chim. Phys.* **24**, 442–459 (1848).
- Sheldon, R. A. in *Chiral Technologies: Industrial Synthesis of Optically Active Compounds* 173–204 (M. Dekker, New York, 1993).
- Addadi, L. & Weiner, S. Crystals, asymmetry and life. *Nature* **411**, 753–754 (2001).
- Jacques, J., Collet, A. & Wilen, S. H. in *Enantiomers, Racemates and Resolution* (Wiley, New York, 1981).
- Maddox, J. Racemization rationalized. *Nature* **341**, 101 (1989).
- De Feyter, S. & De Schryver, F. C. Two-dimensional supramolecular self-assembly probed by scanning tunneling microscopy. *Chem. Soc. Rev.* **32**, 139–150 (2003).
- Ortega Lorenzo, M., Baddeley, C. J., Muryn, C. & Raval, R. Extended surface chirality from supramolecular assemblies of adsorbed chiral molecules. *Nature* **404**, 376–379 (2000).
- Kühnle, A., Linderoth, T. R., Hammer, B. & Besenbacher, F. Chiral recognition in dimerization of adsorbed cysteine observed by scanning tunnelling microscopy. *Nature* **415**, 891–893 (2002).
- Chen, Q. & Richardson, N. V. Enantiomeric interactions between nucleic acid bases and amino acids on solid surfaces. *Nature Mater.* **2**, 324–328 (2003).
- Kuzmenko, I. et al. Aspects of spontaneous separation of enantiomers in two- and three-dimensional crystals. *Chirality* **10**, 415–424 (1998).
- Perez-Garcia, L. & Amabilino, D. B. Spontaneous resolution under supramolecular control. *Chem. Soc. Rev.* **31**, 342–356 (2002).
- De Feyter, S. et al. Homo- and heterochiral supramolecular tapes from achiral, enantiopure, and racemic promesogenic formamides. *Angew. Chem. Int. Edn Engl.* **40**, 3217–3220 (2001).
- Romer, S., Behzadi, B., Fasel, R. & Ernst, K.-H. Homochiral conglomerates and racemic crystals in two dimensions: Tartaric acid on Cu(110). *Chem. Eur. J.* **11**, 4149–4154 (2005).

14. Cai, Y. & Bernasek, S. L. Chiral pair monolayer adsorption of iodine-substituted octadecanol molecules on graphite. *J. Am. Chem. Soc.* **125**, 1655–1659 (2003).
15. Wie, Y., Kannappan, K., Flynn, G. W. & Zimmt, M. B. Scanning tunneling microscopy of prochiral anthracene derivatives on graphite: chain length effects on monolayer morphology. *J. Am. Chem. Soc.* **126**, 5318–5322 (2004).
16. Vidal, F. *et al.* Chiral phase transition in two-dimensional supramolecular assemblies of prochiral molecules. *J. Am. Chem. Soc.* **127**, 10101–10106 (2005).
17. Cai, Y. & Bernasek, S. L. Structures formed by the chiral assembly of racemic mixtures of enantiomers: iodination products of elaidic and oleic acids. *J. Phys. Chem. B* **109**, 4514–4519 (2005).
18. Böhringer, M., Schneider, W.-D. & Berndt, R. Real space observation of a chiral phase transition in a two-dimensional organic layer. *Angew. Chem. Int. Edn Engl.* **39**, 792–795 (2000).
19. Ernst, K.-H., Kuster, Y., Fasel, R., Müller, M. & Ellerbeck, U. Two-dimensional separation of [7]helicene enantiomers on Cu(111). *Chirality* **13**, 675–678 (2001).
20. Parschau, M., Romer, S. & Ernst, K.-H. Induction of homochirality in achiral enantiomorphous monolayers. *J. Am. Chem. Soc.* **126**, 15398–15399 (2004).
21. Fasel, R., Parschau, M. & Ernst, K.-H. Chirality transfer from single molecules into self-assembled monolayers. *Angew. Chem. Int. Edn Engl.* **42**, 5178–5181 (2003).
22. Barlow, S. M. & Raval, R. Complex molecules at metal surfaces: bonding, organisation and chirality. *Surf. Sci. Rep.* **50**, 201–341 (2003).
23. Green, M. M. *et al.* The macromolecular route to chiral amplification. *Angew. Chem. Int. Edn Engl.* **38**, 3138–3154 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Swiss National Science Foundation (SNF). We thank O. Gröning for writing the software to simulate STM images and U. Ellerbeck for the [7]H synthesis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.H.E. (karl-heinz.ernst@empa.ch).

A climatologically significant aerosol longwave indirect effect in the Arctic

Dan Lubin^{1*} & Andrew M. Vogelmann^{2*}

The warming of Arctic climate and decreases in sea ice thickness and extent^{1,2} observed over recent decades are believed to result from increased direct greenhouse gas forcing, changes in atmospheric dynamics having anthropogenic origin^{3–5}, and important positive reinforcements including ice–albedo and cloud–radiation feedbacks⁶. The importance of cloud–radiation interactions is being investigated through advanced instrumentation deployed in the high Arctic since 1997 (refs 7, 8). These studies have established that clouds, via the dominance of longwave radiation, exert a net warming on the Arctic climate system throughout most of the year, except briefly during the summer⁹. The Arctic region also experiences significant periodic influxes of anthropogenic aerosols, which originate from the industrial regions in lower latitudes¹⁰. Here we use multisensor radiometric data^{7,8} to show that enhanced aerosol concentrations alter the microphysical properties of Arctic clouds, in a process known as the ‘first indirect’ effect^{11,12}. Under frequently occurring cloud types we find that this leads to an increase of an average 3.4 watts per square metre in the surface longwave fluxes. This is comparable to a warming effect from established greenhouse gases and implies that the observed longwave enhancement is climatologically significant.

We use five data sets from the North Slope of Alaska (NSA) site, established by US Department of Energy Atmospheric Radiation Measurement (ARM) program⁷, to determine the impact of aerosol on Arctic cloud microphysics and the surface longwave budget. The Atmospheric Emitted Radiance Interferometer (AERI)¹³ is a Fourier transform infrared spectroradiometer that measures downwelling zenith sky radiance in the wavelength range 3.3–25 μm ($400\text{--}3,020\text{ cm}^{-1}$), with a spectral resolution of 1 cm^{-1} and a radiometric calibration accuracy of 1%. The Active Remotely Sensed Cloud Locations (ARSCL)¹⁴ data set combines ceilometer, lidar and millimetre wave cloud radar measurements to determine the vertical location of cloud layers. Downwelling hemispheric longwave flux is measured by pyrgeometers with an accuracy of $\pm 5\%$ (ref. 15). A NOAA Climate Modelling and Diagnostics Laboratory (CMDL) facility¹⁶ is adjacent to the ARM NSA site in Barrow, Alaska, and provides measurements of aerosol condensation nuclei (CN) concentrations at the surface. The CN measurements indicate times when substantial aerosol abundances are present and may have the potential to serve as cloud condensation nuclei in low stratus cloud in the Arctic¹⁷. Rawinsondes are launched approximately five days a week from the NSA site and, in conjunction with the ARSCL data, these soundings provide estimates of cloud base temperature. We analysed six years of the ARM NSA and CMDL data (1998–2003).

AERI spectral measurements in the middle infrared are sensitive to cloud droplet size^{18–20}, and therefore to the aerosol indirect effect. Radiative transfer calculations¹⁹ (Fig. 1a) illustrate how the zenith radiance under a cloud with fixed liquid water path, LWP, varies with

the effective radius of the droplet size distribution^{18,20}, r_e . As r_e decreases, the effective brightness temperature (T_b) increases in the wavenumber (ν) interval $800\text{--}1,000\text{ cm}^{-1}$, and the slope of the spectral brightness temperature, $dT_b/d\nu$, decreases (becomes more negative). This results from changes in cloud emissivity (Fig. 1b); as r_e decreases, emissivity generally increases and, also, its spectral dependence increases within the $800\text{--}1,000\text{ cm}^{-1}$ spectral interval. Figure 1c illustrates the effect of these spectral cloud properties on the downwelling longwave flux at the Arctic Earth surface. For a fixed LWP, a decrease in r_e yields a smaller $dT_b/d\nu$ and a larger downwelling longwave flux. For large LWP, the cloud radiates as a blackbody and the spectral and longwave sensitivities vanish¹⁸. Our approach here is to: (1) detect the aerosol indirect effect by observing whether AERI spectral brightness temperatures indicate a significant shift towards smaller (more negative) $dT_b/d\nu$ in the presence of high CN concentrations; (2) verify a concomitant increase in cloud emissivity, manifested as a larger downwelling flux measured in the pyrgeometer data; (3) quantify the importance of the magnitude of this enhancement to the surface longwave radiation budget; and (4) determine the portion of this longwave enhancement that is due to systematic changes in r_e .

Our analysis indicates a clear tendency for AERI $dT_b/d\nu$ to decrease in the presence of larger CN, consistent with smaller r_e and the aerosol indirect effect. Our approach is illustrated in Fig. 2, which shows two AERI spectra observed under clouds with the same geometrical properties, but which have greatly differing CN concentrations. Consistent with our theoretical discussion, the measured $dT_b/d\nu$ between 800 and $1,000\text{ cm}^{-1}$ is clearly more negative for the high CN case. From this spectral information, it is also possible to directly retrieve r_e and LWP from AERI data^{18,21}, because the spectral radiance in the interval $800\text{--}1,000\text{ cm}^{-1}$ is sensitive to r_e , while the spectral radiance in the interval $1,100\text{--}1,250\text{ cm}^{-1}$ is sensitive to LWP/r_e . From the examples of Fig. 2, we retrieve $\text{LWP} = 15.4\text{ g m}^{-2}$ and $r_e = 11\text{ }\mu\text{m}$ for the low CN case, and $\text{LWP} = 13.5\text{ g m}^{-2}$ and $r_e = 7\text{ }\mu\text{m}$ for the high CN case.

Our detection technique involving $dT_b/d\nu$ was applied to the full six-year data set by collocating individual AERI emission spectra with hourly-averaged CN measurements. The ARSCL data set was used to select all periods when single-layer clouds with base heights and geometric thickness under $1,000\text{ m}$ occurred over surface temperatures between 265 and 280 K . These selection rules favour liquid water clouds. Liquid water was recently discovered to largely govern Arctic cloud radiative properties during spring and summer²², with liquid water being found in clouds at temperatures as low as $-34\text{ }^\circ\text{C}$ (ref. 23). From the AERI measurements, we determined the brightness temperature slope in the interval $800\text{--}1,000\text{ cm}^{-1}$, $dT_b/d\nu$. The CN measurements were screened by wind direction to remove local effects from Barrow. The slopes were sorted by the coincident CN concentration, where ‘low’ CN are $<50\text{ cm}^{-3}$ and ‘high’ are

¹Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0221, USA. ²Brookhaven National Laboratory, Upton, New York 11973, USA.

*These authors contributed equally to this work.

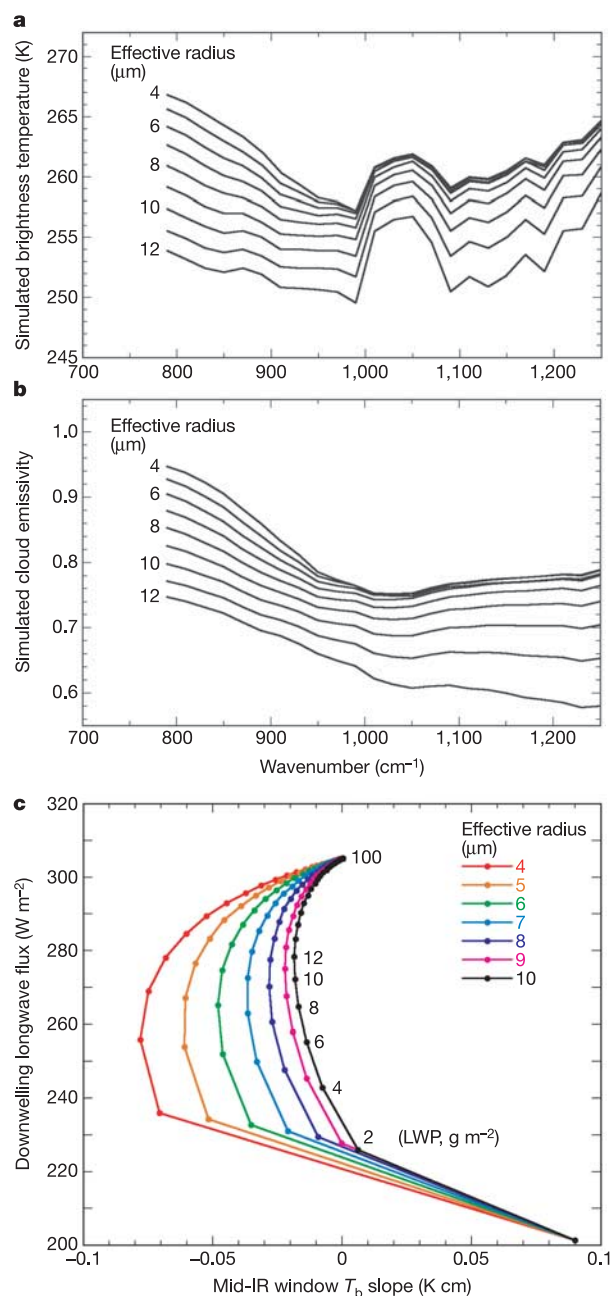


Figure 1 | Demonstration of how surface longwave flux under cloud depends on cloud liquid water path and effective radius. **a**, Radiative transfer simulations¹⁹ of the spectral dependence of effective scene temperature, or 'brightness temperature' (T_b), corresponding to downwelling zenith radiance under a liquid water cloud. The cloud liquid water path (LWP) is held constant at 12 g m^{-2} and the effective radius (r_e) is varied from 4 to $12 \mu\text{m}$, a range commonly observed in terrestrial water clouds. **b**, The spectral cloud emissivity from the calculation in **a**. **c**, Radiative transfer simulation of the downwelling broadband longwave flux as a function of the brightness temperature slope $dT_b/d\nu$ in the interval $800\text{--}1,000 \text{ cm}^{-1}$, plotted as curves of fixed r_e with LWP increasing as indicated.

$>175 \text{ cm}^{-3}$. These thresholds were chosen on the basis of the lowest and highest 25th percentiles for all CN data, respectively, and they are also consistent with other Arctic studies^{17,24}. From the six-year data set subsampled as described, we found 2,379 collocated low CN cases and 5,164 high CN cases. Analysing these cases revealed a systematic tendency for $dT_b/d\nu$ to be more negative under high CN than under low CN, consistent with the aerosol indirect effect causing smaller r_e .

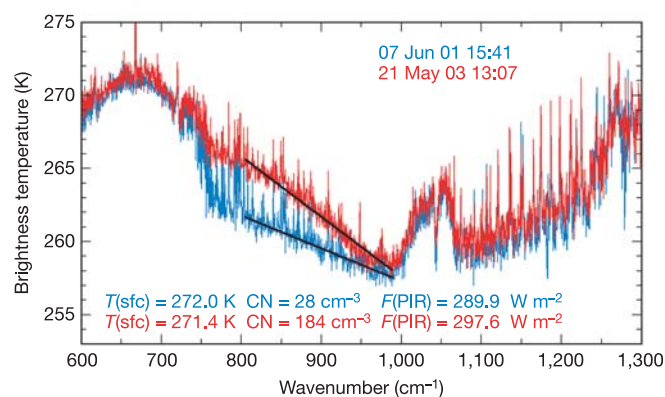


Figure 2 | Examples of AERI measurements. Downwelling emission spectra measured by the NSA AERI beneath two clouds with very different condensation nuclei (CN) concentrations. Near-surface (2 m) air temperature $T(\text{sfc})$ and pyrgeometer-measured downwelling longwave flux $F(\text{PIR})$ are also indicated.

The effects of elevated CN concentrations on the downwelling longwave radiation are investigated by collocating the instantaneous pyrgeometer fluxes with the AERI-measured $dT_b/d\nu$ (Fig. 3a); the relationship between these measured quantities is highly consistent with theoretical expectations (Fig. 1c). For $dT_b/d\nu < 0$, the high CN points clearly exhibit a greater spread towards more negative slopes, and their maximum downwelling pyrgeometer fluxes are greater than for the low CN points. Larger flux for higher CN is consistent with the theorized increase in cloud emissivity for smaller r_e .

This result is better quantified by binning the pyrgeometer fluxes in $dT_b/d\nu$ bins of width 0.01 K cm^{-1} (0.005 K cm^{-1} near $dT_b/d\nu = 0$) and computing the mean and standard deviations of the fluxes per bin. Two-sample t -tests are performed for each $dT_b/d\nu$ bin to determine whether the high and low CN cases comprise statistically significant different distributions. Figure 3b shows that, for the same $dT_b/d\nu$ bin, the pyrgeometer fluxes for high CN are consistently greater than for low CN. The statistical significance is above the 99% confidence level for all but one of the $dT_b/d\nu$ bins. For the $dT_b/d\nu = 0$ bin (rightmost pair of data points), no significant difference exists between the high and low CN cases, consistent with most of those clouds having optical depth greater than ~ 10 and radiating as blackbodies¹⁸. Considering all six years of the low, single-layer stratiform clouds at the NSA, we find that the mean pyrgeometer downwelling flux measured for the high CN cases is 8.2 W m^{-2} greater than for the low CN cases.

We now show that our observed flux difference between the high and low CN cases (Fig. 3b) is not an artefact of additional seasonal considerations or temperature effects. We examine a subset of the springtime data (May and June), for which cloud optical depths have been found to be similar in a recent climatological study²⁵, and for surface temperatures confined to the range $267\text{--}274 \text{ K}$ (in addition to the above selection rules). This subset contains 1,068 collocated low CN and 1,289 high CN cases. Available rawinsonde data covering this data subset indicate average near-surface air temperatures of $272.4 \pm 1.2 \text{ K}$ and $272.3 \pm 1.9 \text{ K}$ for the low and high CN cases, respectively, and cloud base temperatures of $270.4 \pm 3.3 \text{ K}$ and $269.6 \pm 9.1 \text{ K}$ for the low and high CN cases, respectively. We first sorted this data subset on $dT_b/d\nu$ (Fig. 3c), and determined that the mean pyrgeometer downwelling flux measured for these high CN cases is 12.3 W m^{-2} greater than for the corresponding low CN cases, when we exclude the blackbody cases in the $dT_b/d\nu = 0$ bin. We then retrieved r_e and LWP from all AERI spectra having $dT_b/d\nu < 0$ in this data subset^{18,21}. Figure 3d indicates a most probable r_e of $10 \mu\text{m}$ under low CN and $6 \mu\text{m}$ under high CN.

To determine the amount of the total flux difference between these low and high CN cases that is due to this systematic difference in r_e ,

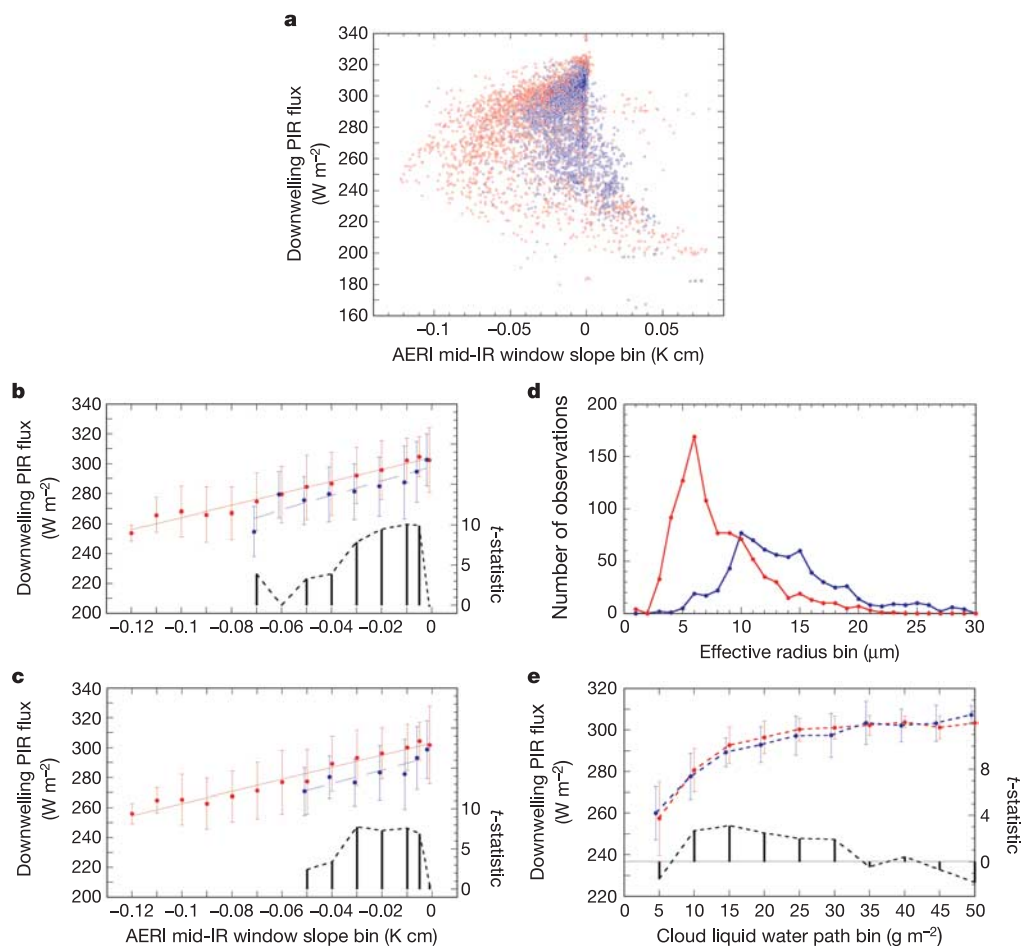


Figure 3 | Demonstration of the aerosol longwave indirect effect from ARM NSA data. **a–c**, The downwelling longwave flux measured by NSA pyrgeometers (Precision Infrared Radiometer, PIR) under single-layer clouds as a function of the AERI-measured dT_b/dv : **a**, instantaneous measurements for all low and high CN cases; **b**, downwelling flux averages and standard deviations (1σ) in dT_b/dv bins of width 0.01 K cm (width 0.005 near $dT_b/dv = 0$) for all low and high CN cases; **c**, as in **b** but for a data

subset comprising May and June only, and surface temperature range 267–274 K. **d**, Histograms of AERI-retrieved r_e under low and high CN from the May–June subset. **e**, Pyrgeometer bin averages as in **c** for the May–June subset, but sorted on LWP bins of width 5 g m^{-2} . Statistics from 2-sample t -tests indicate the bins where significant differences exist between the contrasted sample means for the high and low CN concentrations.

we sorted the May–June data subset in LWP bins of width 5 g m^{-2} (Fig. 3e). Statistically significant flux differences emerge throughout most of the LWP range for which $dT_b/dv < 0$. With this constant-LWP sorting, the mean pyrgeometer flux in a given LWP bin under high CN is on average 3.4 W m^{-2} larger than under low CN. The remaining flux difference between the cases might be an effect of greater LWP under high CN. An increase in cloud LWP in the presence of increased aerosol concentrations would be consistent with recent theoretical work on Arctic cloud microphysics^{26,27}; however, the current analysis does not allow a rigorous evaluation of atmospheric states that would enable us to firmly demonstrate this additional effect.

In conclusion, we provide observational evidence that the first aerosol indirect effect operates in low, optically thin, single-layered Arctic clouds with a concomitant increase in the downwelling longwave flux. The cloud amount during the Arctic spring generally exceeds 80% (ref. 9), which implies that the observed longwave enhancement has climatological significance.

Received 21 June; accepted 14 November 2005.

- Moritz, R. E., Bitz, C. M. & Steig, E. J. Dynamics of recent climate change in the Arctic. *Science* **297**, 1497–1502 (2002).
- Cavaliere, D. J., Gloersen, P., Parkinson, C. L., Comiso, J. C. & Zwally, H. J. Observed hemispheric asymmetry in global sea ice changes. *Science* **278**, 1104–1106 (1997).

- Thompson, D. W. J. & Wallace, J. M. Annular modes in the extratropical circulation. Part I: month-to-month variability. *J. Clim.* **13**, 1000–1016 (2001).
- Rigor, I. G., Wallace, J. W. & Colony, R. L. Response of sea ice to the Arctic Oscillation. *J. Clim.* **15**, 2648–2663 (2002).
- Johannessen, O. M. *et al.* Arctic climate change: observed and modelled temperature and sea ice variability. *Tellus A* **56**, 328–341 (2004).
- Curry, J. A. & Webster, P. J. *Thermodynamics of Atmospheres and Oceans* (Academic, San Diego, 1999).
- Stamnes, K., Ellingson, R. G., Curry, J. A., Walsh, J. E. & Zak, B. D. Review of science issues, deployment strategy, and status for the ARM North Slope of Alaska–Adjacent Arctic Ocean climate research site. *J. Clim.* **12**, 46–63 (1999).
- Uttal, T. *et al.* Surface heat budget of the Arctic Ocean. *Bull. Am. Meteorol. Soc.* **83**, 255–275 (2002).
- Intrieri, J. M., Shupe, M. D., Uttal, T. & McCarty, B. J. An annual cycle of Arctic cloud characteristics observed by radar and lidar at SHEBA. *J. Geophys. Res.* **107**, doi:10.1029/2000JC000423 (2002).
- Barrie, L. A. Arctic air pollution: an overview of current knowledge. *Atmos. Environ.* **20**, 643–663 (1986).
- Twomey, S. The influence of pollution on the shortwave albedo of clouds. *J. Atmos. Sci.* **34**, 1149–1152 (1977).
- Garrett, T., Radke, L. F. & Hobbs, P. V. Aerosol effects on cloud emissivity and surface longwave heating in the Arctic. *J. Atmos. Sci.* **59**, 769–778 (2002).
- Knuteson, R. O. *et al.* Atmospheric Emitted Radiance Interferometer. Part II: instrument performance. *J. Ocean. Atmos. Technol.* **21**, 1777–1789 (2004).
- Clothiaux, E. E. *et al.* Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. *J. Appl. Meteorol.* **39**, 645–665 (2000).
- Payne, R. E. & Anderson, S. P. A new look at calibration and use of Eppley Precision Infrared Radiometer, Part II: calibration and use of the Woods Hole

- Oceanographic Institution improved meteorology precision infrared radiometer. *J. Ocean. Atmos. Technol.* **16**, 739–751 (1999).
16. Delene, D. J. & Ogren, J. A. Variability of aerosol optical properties at four North American surface monitoring sites. *J. Atmos. Sci.* **59**, 1135–1150 (2002).
 17. Garrett, T. J., Zhao, C., Dong, X., Mace, G. G. & Hobbs, P. V. Effects of varying aerosol regimes on low-level Arctic stratus. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL019928 (2004).
 18. Turner, D. D. Arctic mixed-phase cloud properties from AERI lidar observations: Algorithm and results from SEHBA. *J. Appl. Meteorol.* **44**, 427–444 (2005).
 19. Lubin, D. Thermodynamic phase of maritime Antarctic clouds from FTIR and supplementary radiometric data. *J. Geophys. Res.* **109**, doi:10.1029/2003JD003979 (2004).
 20. Mahesh, A., Walden, V. P. & Warren, S. G. Ground-based infrared remote sensing of cloud properties over the Antarctic Plateau. Part II: cloud optical depths and particle sizes. *J. Appl. Meteorol.* **40**, 1279–1294 (2001).
 21. Guo, G., Ji, Q., Yang, P. & Tsay, S.-C. Remote sensing of cirrus optical and microphysical properties from ground-based infrared radiometric measurements—Part II: Retrievals from CRYSTAL-FACE measurements. *IEEE Geosci. Rem. Sens. Lett.* **2**, 132–135 (2005).
 22. Shupe, M. D. & Intrieri, J. M. Cloud radiative forcing of the Arctic surface: the influence of cloud properties, surface albedo, and solar zenith angle. *J. Clim.* **17**, 616–628 (2004).
 23. Intrieri, J. M. et al. An annual cycle of surface cloud forcing at SHEBA. *J. Geophys. Res.* **107**, doi:10.1029/2000JC000439 (2002).
 24. Penner, J. E., Dong, X. & Chen, Y. Observational evidence of a change in radiative forcing due to the aerosol indirect effect. *Nature* **427**, 231–234 (2004).
 25. Leontyeva, E. & Stamnes, K. Estimations of cloud optical thickness from ground-based measurements of incoming solar radiation in the Arctic. *J. Clim.* **7**, 566–578 (1994).
 26. Jiang, H. J., Feingold, G., Cotton, W. R. & Duynkerke, P. G. Large-eddy simulations of entrainment of cloud condensation nuclei into the Arctic boundary layer: May 18, 1998, FIRE/SHEBA case study. *J. Geophys. Res.* **106**, 15113–15122 (2001).
 27. Morrison, H., Curry, J. A., Shupe, M. D. & Zuidema, P. A new double-moment microphysics parameterization for application in cloud and climate models. Part II: Single-column modeling of arctic clouds. *J. Atmos. Sci.* **62**, 1678–1693 (2005).

Acknowledgements This work was supported by the DOE ARM programme. We thank J. Ogren for access to the NOAA CMDL aerosol data.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.L. (dlubin@ucsd.edu).

Universal scaling of respiratory metabolism, size and nitrogen in plants

Peter B. Reich¹, Mark G. Tjoelker², Jose-Luis Machado³ & Jacek Oleksyn⁴

The scaling of respiratory metabolism to body size in animals is considered to be a fundamental law of nature^{1–11}, and there is substantial evidence for an approximate $\frac{3}{4}$ -power relation. Studies suggest that plant respiratory metabolism also scales as the $\frac{3}{4}$ -power of mass^{12–14}, and that higher plant and animal scaling follow similar rules owing to the predominance of fractal-like transport networks and associated allometric scaling^{8–14}. Here, however, using data obtained from about 500 laboratory and field-grown plants from 43 species and four experiments, we show that whole-plant respiration rate scales approximately isometrically (scaling exponent ≈ 1) with total plant mass in individual experiments and has no common relation across all data. Moreover, consistent with theories about biochemically based physiological scaling^{15–18}, isometric scaling of whole-plant respiration rate to total nitrogen content is observed within and across all data sets, with a single relation common to all data. This isometric scaling is unaffected by growth conditions including variation in light, nitrogen availability, temperature and atmospheric CO₂ concentration, and is similar within or among species or functional groups. These findings suggest that plants and animals follow different metabolic scaling relations, driven by distinct mechanisms.

Abundant evidence indicates that diverse fundamental characteristics of animals and plants, including physiological, morphological and allometric attributes, scale with increasing size^{1–14}, as described by power laws of the form:

$$Y = Y_0 M^b \quad (1)$$

where Y is an attribute such as metabolic rate, Y_0 is a normalization constant, M is body mass, and b is the scaling exponent. Perhaps the most fundamental of all such relationships is the one relating whole-organism metabolic rates of animals as the $\sim \frac{3}{4}$ -power of body size^{1–4}; in other words, with each 10-fold increase in size, metabolic rate increases by 7.5-fold. This has been shown to be true for basal metabolism and field metabolism in animals, and for numerous endothermic and ectothermic organisms^{1–4}. Although it was originally expected that these allometric scaling laws would follow a $\frac{2}{3}$ exponent because of the Euclidean surface-area rule^{2–4}, the preponderance of data^{1–4} suggest that the exponent is $\frac{3}{4}$, although vigorous debate continues on this point^{5–7}.

Some argue that $\frac{3}{4}$ -power allometric scaling in biology is so pervasive that various allometric relations probably have a common, mechanistic origin^{4,8,9}. Thus, there has been a surge in theoretical work aimed at explaining scaling relations^{8–11}, as well as debate about the predictive value and validity of such models^{5–7,19}. The disproportionate increase (less than isometric) in metabolism with increasing size has been explained by various metabolic scaling theories^{8–11}, including one based on the fractal-like design of exchange surfaces

and space-filling resource distribution networks, such as the animal vascular system. This theory suggests that physiology is constrained by hierarchical branching networks that optimally supply resources to all parts of three-dimensional organisms, and that such constraints are similar for vascular plants as they are for animals^{8,9}. Evidence indicates that allometric and physiological scaling in plants often follow $\frac{3}{4}$ -power laws similar to those found for animals^{12,20–25}. A hierarchically branched model of plant geometry and hydraulics predicted that whole-plant metabolism would scale to the $\frac{3}{4}$ -power with mass, but there were no empirical data available to test this particular prediction¹².

A subsequent model, based on published metabolic (that is, basal respiration for animals, dark respiration for plants) rates of diverse taxa, predicted a common $\frac{3}{4}$ -power size-dependent scaling (as well as common temperature-dependent scaling) for unicellular organisms, plants and animals¹³. This ‘universal’ relationship has been developed in various forms in subsequent publications^{14,26}, further contributing to the notion of common size-dependent scaling in plants and animals. However, the plant data used in establishing these models and estimates^{13,14} were sparse and disparate: they included observations of 20 different individual plants or plant parts, including fruits, seeds and storage organs, and three small individuals of one species, measured at different temperatures and sizes¹³. Thus, a more comprehensive test of predicted metabolism–size scaling relations in plants is needed that encompasses a wider array of taxa and environmental conditions.

To address this information gap and to test both prior theoretical predictions^{12,13} and hypotheses about the biochemical control of the scaling of plant respiration^{15–18}, here we present data for ~ 500 observations from 43 perennial plant species of coupled measurements of whole-plant dry mass, nitrogen (N) content and respiration rate from four separate studies of laboratory and wild field-grown plants that ranged in age from 1 month to 25 yr (see Methods and Supplementary Information). The data include a wide range of species and functional groups, and plants were grown under a heterogeneous set of environmental conditions that included experimentally controlled contrasts involving temperature, light, N supply and atmospheric CO₂ concentration. Moreover, we assess above-ground data separately to enable the inclusion of published above-ground respiration data for an additional ten large trees^{27,28} to extend the size range of our analyses. Collectively, the observations span five of the roughly 12 orders of magnitude of size in vascular plants²².

The results of each experiment do not support the idea of $\frac{3}{4}$ -power scaling and instead are consistent with approximately isometric (exponent ≈ 1.0) scaling of respiration to body mass for whole plants (Tables 1 and 2, and Fig. 1a). For the four independent studies, the mean scaling exponent of respiration to mass was 1.04, and two-tailed t -tests show the average exponent among the four studies to be

¹Department of Forest Resources, University of Minnesota, St Paul, Minnesota 55108, USA. ²Department of Forest Science, Texas A&M University, College Station, Texas 77843, USA. ³Department of Biology, Swarthmore College, Swarthmore, Pennsylvania 19081, USA. ⁴Polish Academy of Sciences, Institute of Dendrology, PL-62-035 Kornik, Poland.

Table 1 | Scaling of whole-plant respiration and dry mass by experiment and environment

Study	Treatment or growth condition	<i>n</i>	Intercept	Exponent	Lower CI	Upper CI	<i>r</i>
Field/US tree saplings	All	119	0.391	1.117	1.052	1.185	0.951
GH/tree seedlings	All	165	1.345	1.037	1.002	1.073	0.976
GC/tree seedlings	All	190	1.472	1.058	1.021	1.097	0.970
GH/herb seedlings	All	62	1.433	0.967	0.911	1.026	0.974
Lab/all	All indoor seedlings	417	1.426	1.047	1.023	1.073	0.970
Field/US tree saplings	Understory	77	0.389	1.110	1.017	1.190	0.947
Field/US tree saplings	Small gap	42	0.426	1.120	1.017	1.233	0.958
GH/tree seedlings	Shaded	39	1.546	1.120	1.022	1.226	0.965
GH/tree seedlings	Unshaded	126	1.323	1.058	1.020	1.097	0.979
GC/tree seedlings	Ambient CO ₂	96	1.476	1.058	1.003	1.115	0.969
GC/tree seedlings	Increased CO ₂	94	1.468	1.060	1.008	1.114	0.972
GC/tree seedlings	18/12 °C	63	1.549	1.000	0.935	1.070	0.968
GC/tree seedlings	24/18 °C	64	1.469	1.050	0.996	1.107	0.979
GC/tree seedlings	30/24 °C	63	1.382	1.110	1.060	1.163	0.984
GH/herb seedlings	Ambient N	31	1.335	0.925	0.849	1.008	0.976
GH/herb seedlings	Added N	31	1.464	0.928	0.865	0.996	0.983

All equations were fitted by the log-log version of the equation: $Y = Y_0 M^b$. Reduced major axis intercepts and slopes (exponents) are shown, as well as the lower and upper 95% CI of the exponent and the Pearson correlation coefficient (*r*). Field/US tree saplings, field study of saplings of four tree species; GH/tree seedlings, greenhouse study of seedlings of nine tree species; GC/tree seedlings, growth chamber study of seedlings of five tree species; GH/herb seedlings, greenhouse study of seedlings of 32 herbaceous species; *n*, number of observations.

significantly different from 0.75 ($P = 0.002$) but not from 1.0 ($P = 0.244$).

A robust test of scaling relationships is whether they are consistent within narrow and broad comparative groupings by growth environment or taxa. In each study, relations for plants in contrasting environments (high versus low light, differing atmospheric CO₂ concentration, temperature or N supply) or different functional groups had similar scaling slopes, with the position of the lines (that is, the intercepts) showing greater variability (Tables 1 and 2). Among our three studies involving woody plants, we could evaluate the whole-plant metabolism versus mass scaling relationship at the species level for 16 cases, each representing a different individual species by experiment combination (ten species, some repeated in more than one study). For those 16 relationships (Table 2), the mean scaling exponent was 1.03 (95% confidence interval (CI) = 0.94–1.11), which was significantly different from 0.75 ($P < 0.0001$), but not from 1.0 ($P = 0.526$).

A common metabolism–size relationship fitted all first-year greenhouse and growth chamber plants pooled ($n = 417$, scaling exponent = 1.05; Table 1 and Fig. 1a), even though the data included tree, grass and forb species grown under differing environmental conditions in distinct experiments. Older, field-grown plants (US tree saplings) had a scaling slope of respiration to plant mass similar

to that of the greenhouse and growth chamber plants. However, the field-grown plants had an intercept of this relation that was markedly different to that of the laboratory plants (Fig. 1a); thus, whole-plant respiration rates at any given size were lower for the field-grown plants than the laboratory plants (Fig. 1a). In addition, both the field-grown saplings (US tree saplings) and larger trees (Japan trees) had much greater above-ground biomass at a common above-ground respiration rate than did the laboratory-grown plants (Fig. 1d).

Fitting a relationship of respiration to mass across all available data yields scaling exponents of 0.81 (95% CI = 0.79–0.84) and 0.84 (95% CI = 0.82–0.86) for the whole-plant and above-ground data sets, respectively. However, plots of residuals versus predicted values show non-constant residual mean and residual variance functions, indicating that a single overall model fit is inappropriate. This is because these correlations are fitted across data sets with similar slopes but different intercepts (Fig. 1a, d). It is therefore difficult to quantify a comprehensive relationship of respiration to plant mass across all plant sizes and growth environments, although a scaling exponent of <1 across the full range of plant sizes is consistent with the widely held notion that the fraction of tissues with low respiration rate (such as boles of large trees) increases with plant size. However, even ignoring statistical concerns, which could be a result of the nature of our data compilations, a different slope across all data (~ 0.81 –0.84)

Table 2 | Scaling of whole-plant respiration and dry mass by species and functional group

Study	Group or Species	<i>n</i>	Intercept	Exponent	Lower CI	Upper CI	<i>r</i>
GH/herb seedlings	Grasses	32	1.424	0.961	0.860	1.074	0.959
GH/herb seedlings	Forbs	30	1.444	0.973	0.915	1.036	0.987
Field/US tree saplings	<i>Abies balsamifera</i>	35	0.019	1.293	1.198	1.394	0.978
Field/US tree saplings	<i>Acer rubrum</i>	35	0.277	1.189	1.059	1.336	0.951
Field/US tree saplings	<i>Betula papyrifera</i>	14	1.108	0.714	0.631	0.807	0.982
Field/US tree saplings	<i>Pinus strobus</i>	35	0.250	1.183	1.091	1.283	0.975
GH/tree seedlings	<i>Betula alleghaniensis</i>	9	1.379	0.923	0.827	1.030	0.998
GH/tree seedlings	<i>Betula papyrifera</i>	35	1.374	0.934	0.889	0.981	0.991
GH/tree seedlings	<i>Larix laricina</i>	18	1.402	1.017	0.921	1.122	0.983
GH/tree seedlings	<i>Picea glauca</i>	13	1.428	1.233	0.920	1.652	0.919
GH/tree seedlings	<i>Pinus banksiana</i>	19	1.436	1.110	0.863	1.428	0.901
GH/tree seedlings	<i>Pinus strobus</i>	23	1.373	1.237	1.099	1.391	0.968
GH/tree seedlings	<i>Populus tremuloides</i>	34	1.397	0.917	0.873	0.963	0.991
GC/tree seedlings	<i>Betula papyrifera</i>	36	1.512	0.882	0.834	0.932	0.988
GC/tree seedlings	<i>Larix laricina</i>	40	1.359	0.927	0.874	0.984	0.984
GC/tree seedlings	<i>Picea mariana</i>	42	1.284	0.967	0.888	1.053	0.967
GC/tree seedlings	<i>Pinus banksiana</i>	42	1.325	0.990	0.923	1.062	0.977
GC/tree seedlings	<i>Populus tremuloides</i>	30	1.714	0.906	0.853	0.963	0.988

The studies, analyses and abbreviations are as in Table 1. Two species in the GH/tree seedling study had either insufficient sample size or a non-significant relation and are not included.

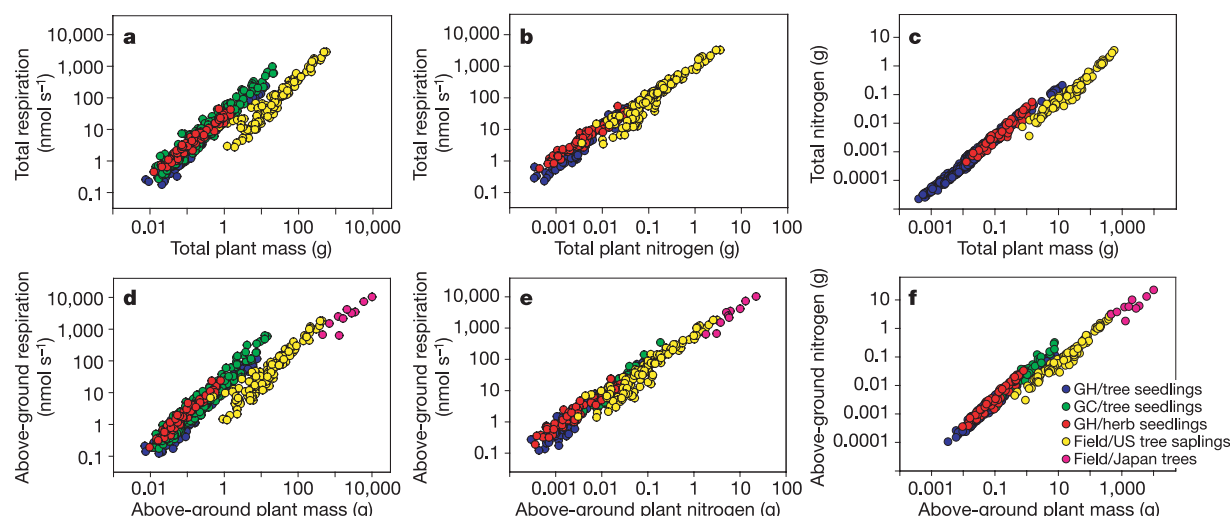


Figure 1 | Scaling of respiration, N and plant mass for plants. Data are from studies of field, greenhouse (GH) and growth chamber (GC) plants.

a, Whole-plant respiration in relation to total plant dry mass. **b**, Whole-plant respiration in relation to whole-plant N. **c**, Whole-plant N in relation to total plant dry mass. **d**, Above-ground plant respiration in relation to

above-ground plant dry mass. **e**, Above-ground plant respiration in relation to above-ground plant N. **f**, Above-ground plant N in relation to above-ground plant dry mass. Respiration, determined as net CO₂ efflux, was adjusted to a common measurement temperature (see Methods and Supplementary Information). See Tables 1–3 for details.

from that within data sets (~ 1.00 – 1.05) is inconsistent with the idea of a general relation involving size *per se*.

In contrast to the lack of a single universal relationship between plant respiration and plant mass, the data for all plants from all studies, including field and laboratory, are described by a single, common relationship between total respiration and total plant N content (Fig. 1b) or above-ground respiration and above-ground N content (Fig. 1e). The scaling exponents of these relationships (Table 3 and Fig. 1) are nearly 1.0 (for example, for total plant data, exponent = 1.03, 95% CI = 1.00–1.05). Thus, the relationship between respiration and N content reconciles intercept differences among studies observed for the respiration versus plant mass relationships. These results are consistent with the fundamental role of N in the biochemistry of plant respiration and strong evidence for universal relations between respiration and N in higher plants^{15–18}.

Total plant N also scaled nearly isometrically with total dry mass in our individual studies (average exponent ≈ 0.98 for the four studies; Table 3 and Fig. 1c). Similar to respiration, however, field-grown tree saplings in a Minnesota forest and plantation-grown trees in Japan had lower plant N content at a given plant mass (that is, they had

lower tissue N concentration and a lower elevation of the scaling relation) than did first-year plants grown in the greenhouse or growth chambers (Fig. 1c, f). Fitting a single relationship across all data, total plant N scales as the 0.81-power of total mass (95% CI = 0.80–0.82), but plots of residuals versus predicted values again indicate that a single overall model fit is inappropriate. Similarly, across all data, plant N per unit mass (that is, N concentration) declines significantly ($P < 0.05$) with increasing plant mass, although this relationship is a result of a clustering of clouds of data points (representing individual studies) in different positions along the axes, and is significant in only two of our four individual data sets.

The mechanistic coupling of plant respiration and N (refs 15–18), the universal scaling of respiration and N (Fig. 1b, e), and variability in plant N concentration provide a physiologically based explanation, first, for isometric scaling of respiration to mass in studies that encompass both ontogenetic and interspecific variation in contrasting growth environments; second, for differences in the intercept of this relationship among studies; and third, for the non-isometric scaling of respiration versus mass across all data pooled. In effect, the lower intercept (elevation) of the respiration–mass scaling relation of the wild field-grown saplings in Minnesota or plantation trees in Japan as compared with the laboratory-grown seedlings (Fig. 1a, d) can be explained by their lower N content or concentration at a given plant size (Fig. 1c, f).

The results of our study of respiration and body size in whole plants provide no support for a key prediction of the theory of $\frac{3}{4}$ -power scaling of metabolism to body mass. Instead, these data indicate that there is no universal, fixed scaling of respiration versus size in plants, because individual studies have similar, near isometric (exponent ≈ 1.04) scaling within data sets, but can differ in intercept, resulting in a (statistically problematic) scaling exponent of <1.0 but >0.75 across the widest range in plant sizes. By contrast, the only relationship common across all data in our compilation is that relating respiration per plant and N per plant.

Why should plant metabolism scale as ~ 1.0 -power of N and not as $\sim \frac{3}{4}$ -power of mass? First, plant tissues require N as a chief component of key enzymes to carry on crucial metabolic processes. Specific rates of respiration and N concentration in plant leaves and roots are often positively correlated, especially for broad comparisons among species^{15–18}. In our study, when respiration and N are expressed per unit mass, whole-plant specific respiration rate scales positively with

Table 3 | Scaling of plant N with plant respiration and dry mass

Study	<i>n</i>	Intercept	Exponent	Lower CI	Upper CI	<i>r</i>
Plant N versus plant mass						
Field/US tree saplings	118	−2.197	0.945	0.908	0.983	0.977
GH/tree seedlings	460	−1.631	0.920	0.910	0.931	0.993
GC/tree seedlings	46	−1.528	1.060	0.971	1.156	0.961
GH/herb seedlings	63	−1.517	1.003	0.944	1.066	0.973
Total plant respiration versus total plant N						
Field/US tree saplings	118	2.985	1.182	1.114	1.253	0.952
GH/tree seedlings	135	3.185	1.137	1.102	1.173	0.984
GH/herb seedlings	62	2.889	0.960	0.905	1.109	0.975
All	315	2.922	1.028	1.004	1.052	0.979
Above-ground plant respiration versus above-ground plant N						
All	381	2.920	1.048	1.027	1.069	0.981

Shown are data from four experiments (as in Table 1), as well as the relation of above-ground respiration versus N for all data pooled, including additional data points for Japanese trees (see Fig. 1f). For the GC/tree seedling study, the plant N versus plant mass relation is shown for shoot (stem plus foliage) N and mass, because no root N data are available.

N per unit mass (that is, N concentration) within studies and across all data pooled (exponent = 1.17, 95% CI = 1.11–1.24), although the fits are poorer than those for whole plants (data not shown). The strong physiological coupling of respiratory metabolism with N, in combination with an overall range of percentage N that is modest compared with the enormous range of total N content across size, apparently leads to common and roughly isometric respiration–N relations across all plants.

Second, the isometric scaling of respiration rate and N does not lead to a $\frac{3}{4}$ -power scaling of respiration to mass, as it would if N scaled as the $\frac{3}{4}$ -power of mass. Although the $\frac{3}{4}$ -power law scaling in animals has been explained to result from the efficient design of exchange surfaces and vascular distribution networks, the N concentration in any given plant tissue and the total N content in any given plant are not just a passive reflection of vascular networks. The concentration and total pools of N in any plant are in part governed by mechanisms involved in N uptake, initial allocation and subsequent resorption and redistribution. For example, perennial plants resorb roughly half of the N in leaves before they are shed, which supplies a sizeable fraction of future plant N demand. Taxa differ at the leaf and whole-plant level in the magnitude of this proportional N recovery and thus in the fraction of future N demand met by this conserved N. This is one example of an important mechanism in plants that can partially uncouple N content from constraints of the vascular network.

In addition, other aspects of plant respiration are less tightly linked to vascular networks than are similar processes in animals. Plants and animals differ fundamentally in O_2 and CO_2 exchange processes. At the capillaries—the end points of the vascular delivery system in animals—the supply of all principal resources (O_2 , carbon substrates and mineral elements) required for respiratory metabolism is constrained by the geometry of the vascular network. By contrast, in plants the most metabolically active tissues—leaves—are at the distal ends of the delivery network for nutrients and water, but O_2 and CO_2 are exchanged with the atmosphere by diffusion directly into leaves, with an additional set of factors further constraining these fluxes. Furthermore, substrate supply to mitochondrial respiration in plants is linked to carbohydrates produced in photosynthesis, with source–sink relationships governing internal transport, which may also uncouple metabolism from vascular constraints. Thus, respiratory metabolism in plants is linked to N-rich enzymes, substrate supply and adenylate demand^{15–18}, which are not necessarily under dominant control by vascular constraints. Finally, research suggests that there may not be universal allometric scaling of hydraulic architecture and water transport in plants^{29,30}. Consequently, even the supply of resources such as water that logically should be constrained by the plant hydraulic network, may not be as uniformly controlled by that system as has been proposed^{8,12}. Thus, the processes that influence respiration in plants stand in contrast to those in animals where the supply of respiratory substrate and oxygen and CO_2 gas exchange may be more closely constrained by vascular networks.

In summary, we consistently observed near isometric scaling of whole-plant respiration, N content and plant mass in experiments that each included various vascular plant taxa and growing environments, and common scaling of total plant respiration to total plant N across all taxa, environments and experiments. These results are consistent with the notion of convergent scaling of metabolism to tissue N concentration^{16–18}. Moreover, these findings are different than the $\frac{3}{4}$ -power scaling found for metabolism and body size in animals, and suggest that the notion of a single general law of size-dependent metabolism^{12–14} may be premature. Developing general scaling models that can predict scaling relationships for both plants and animals represents an important challenge to biology and ecology.

METHODS

Data compilation. The data used in these analyses were compiled from several studies designed to address questions about environmental effects (for example,

light level, temperature or atmospheric CO_2 concentration) on plant growth, morphology and physiology (Supplementary Information), including whether these varied with plant size, but these data were not previously used to assess broad issues of metabolic scaling among plant taxa or environmental conditions.

Our data compilation for whole plants included coupled measurements of dark respiration rates, plant size and N content for first-year plants grown from seed in laboratory or greenhouse conditions in three separate experiments, and for 6- to 25-yr-old tree saplings growing naturally in a temperate forest. Because all species inhabit the temperate zone, respiration measurements were made during the growing season summer or summer-like conditions in controlled environments. The three seedling studies included GH/tree seedlings: a greenhouse experiment with nine cold temperate tree species grown in shaded and unshaded conditions ($n = 165$ seedlings measured for respiration); GC/tree seedlings: a growth chamber experiment with five cold temperate tree species growing in contrasting atmospheric CO_2 concentrations and in three different temperature regimes ($n = 190$ seedlings); and GH/herbs: a greenhouse experiment with 32 herbaceous grassland species (including grasses and forbs) growing in ambient or N-enriched soil ($n = 62$ seedlings). A fourth study (Field/US tree saplings) involved naturally regenerated saplings of four cold temperate tree species growing in contrasting light microhabitats (understory and small gap) in a forest in northern Minnesota, USA ($n = 119$ seedlings).

Measurements. Dark respiration rates were measured by infrared gas analysis of net CO_2 efflux at temperatures near growth temperatures for the intact foliage, stem (or foliage plus stem), and root systems of individual plants, and then aggregated to the whole plant (see Supplementary Information for details). In addition, to extend the range of plant sizes to encompass large trees (Field/Japan trees), we compiled published above-ground data on night time net CO_2 efflux collected during the summer on ten *Chamaecyparis obtusa* trees^{27,28}. Where necessary, we adjusted all respiration rates to a common measurement temperature (24 °C) by using a published temperature model¹⁵ to reconcile measurement temperature differences among studies. Results are similar using data adjusted to a common temperature, data reported at growth temperature, or data recorded at measurement temperatures (see Supplementary Information for details).

Data analysis. Standardized major axis slopes were fitted to bivariate trait relationships of log-transformed variables using type II linear regression. The slope of such lines represents the proportional (scaling) relationship and is equivalent to the scaling exponent of the power law (equation (1)). We also used analysis of covariance, which is based on type I linear regression, to test for slope differences among groups and, when slopes were similar, for intercept differences in the relations of the log-transformed variables.

Received 20 June; accepted 29 September 2005.

- Kleiber, M. Body size and metabolism. *Hilgardia* **6**, 315–353 (1932).
- Peters, R. H. *The Ecological Implications of Body Size* (Cambridge Univ. Press, Cambridge, MA, 1983).
- Calder, W. A. *Size, Function, and Life History* (Harvard Univ. Press, Cambridge, MA, 1984).
- Savage, V. M. et al. The predominance of quarter-power scaling in biology. *Funct. Ecol.* **18**, 257–282 (2004).
- Riisgård, H. U. No foundation of a '3/4 power scaling law' for respiration in biology. *Ecol. Lett.* **1**, 71–73 (1998).
- Dodds, P. S., Rothman, D. H. & Weitz, J. S. Re-examination of the $\frac{3}{4}$ -law of metabolism. *J. Theor. Biol.* **209**, 9–27 (2001).
- White, C. R. & Seymour, R. S. Mammalian basal metabolic rate is proportional to body mass. *Proc. Natl Acad. Sci. USA* **100**, 4046–4049 (2003).
- West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
- West, G. B., Brown, J. H. & Enquist, B. J. The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* **284**, 1677–1679 (1999).
- Banavar, J. R., Maritan, A. & Rinaldo, A. Size and form in efficient transportation networks. *Nature* **399**, 130–134 (1999).
- Banavar, J. R., Damuth, J., Maritan, A. & Rinaldo, A. Supply–demand balance and metabolic scaling. *Proc. Natl Acad. Sci. USA* **99**, 10506–10509 (2002).
- West, G. B., Brown, J. H. & Enquist, B. J. A general model for the structure, and allometry of plant vascular systems. *Nature* **400**, 664–667 (1999).
- Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M. & Charnov, E. L. Effects of size and temperature on metabolic rate. *Science* **293**, 2248–2251 (2001).
- Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789 (2004).
- Atkin, O. K. & Tjoelker, M. G. Thermal acclimation and the dynamic response of plant respiration to temperature. *Trends Plant Sci.* **8**, 343–351 (2003).
- Ryan, M. G. Foliar maintenance respiration of subalpine and boreal trees and shrubs in relation to nitrogen content. *Plant Cell Environ.* **18**, 765–772 (1995).
- Reich, P. B. et al. Relationships of leaf dark respiration to leaf nitrogen, specific

- leaf area and leaf life-span: a test across biomes and functional groups. *Oecologia* **114**, 471–482 (1998).
18. Wright, I. J. *et al.* The worldwide leaf economics spectrum. *Nature* **428**, 821–827 (2004).
 19. Kozłowski, J. & Konarzewski, M. Is West, Brown and Enquist's model of allometric scaling mathematically correct and mathematically relevant? *Funct. Ecol.* **18**, 283–289 (2004).
 20. Enquist, B. J., Brown, J. H. & West, G. B. Allometric scaling of plant energetics and population density. *Nature* **395**, 163–166 (1998).
 21. Reich, P. B. Body size, geometry, longevity and metabolism: do plant leaves behave like animal bodies? *Trends Ecol. Evol.* **16**, 674–680 (2001).
 22. Enquist, B. J. Universal scaling in tree and vascular plant allometry: towards a general quantitative theory linking plant form and function from cells to ecosystems. *Tree Phys.* **22**, 1045–1064 (2002).
 23. Enquist, B. J. & Niklas, K. J. Global allocation rules for biomass partitioning in seed plants. *Science* **295**, 1517–1520 (2002).
 24. Niklas, K. J. & Enquist, B. J. On the vegetative biomass partitioning of seed plant leaves, stems, and roots. *Am. Nat.* **159**, 482–497 (2002).
 25. Niklas, K. J. & Enquist, B. J. Invariant scaling relationships for interspecific plant biomass production rates and body size. *Proc. Natl Acad. Sci. USA* **98**, 2922–2927 (2001).
 26. Enquist, B. J. *et al.* Scaling metabolism from organisms to ecosystems. *Nature* **423**, 639–642 (2003).
 27. Yokota, T., Ogawa, K. & Hagihara, A. Dependence of the above-ground respiration of hinoki cypress (*Chamaecyparis obtusa*) on tree size. *Tree Physiol.* **14**, 467–479 (1994).
 28. Yokota, T. & Hagihara, A. Changes in the relationship between tree size and aboveground respiration in field-grown hinoki cypress (*Chamaecyparis obtusa*) trees over three years. *Tree Physiol.* **18**, 37–43 (1998).
 29. Meinzer, F. C., Bond, B. J., Warren, J. M. & Woodruff, D. R. Does water transport scale universally with tree size? *Funct. Ecol.* **19**, 558–565 (2005).
 30. McCulloh, K. A. & Sperry, J. S. Patterns in hydraulic architecture and their implications for transport efficiency. *Tree Physiol.* **25**, 257–267 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Wilderness Research Foundation and the National Science Foundation Long-Term Ecological Research Program for major support of the research reported herein and the Gordon Conference 'Metabolic Basis of Ecology' for inspiring us in this endeavour.

Author Contributions The idea behind this paper was developed by P.B.R. and J.-L.M. while attending the Gordon Conference *Metabolic Basis of Ecology*. P.B.R. did the statistical analyses and wrote the paper with considerable assistance from M.G.T. All authors interpreted the results, commented on the manuscript and were involved in the design and implementation of one or more of the individual studies that collectively make up this paper (P.B.R., four studies; M.G.T., three; J.-L.M., two; J.O., one).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.B.R. (preich@umn.edu).

LETTERS

The scaling laws of human travel

D. Brockmann^{1,2}, L. Hufnagel³ & T. Geisel^{1,2,4}

The dynamic spatial redistribution of individuals is a key driving force of various spatiotemporal phenomena on geographical scales. It can synchronize populations of interacting species, stabilize them, and diversify gene pools^{1–3}. Human travel, for example, is responsible for the geographical spread of human infectious disease^{4–9}. In the light of increasing international trade, intensified human mobility and the imminent threat of an influenza A epidemic¹⁰, the knowledge of dynamical and statistical properties of human travel is of fundamental importance. Despite its crucial role, a quantitative assessment of these properties on geographical scales remains elusive, and the assumption that humans disperse diffusively still prevails in models. Here we report on a solid and quantitative assessment of human travelling statistics by analysing the circulation of bank notes in the United States. Using a comprehensive data set of over a million individual displacements, we find that dispersal is anomalous in two ways. First, the distribution of travelling distances decays as a power law, indicating that trajectories of bank notes are reminiscent of scale-free random walks known as Lévy flights. Second, the probability of remaining in a small, spatially confined region for a time T is dominated by algebraically long tails that attenuate the superdiffusive spread. We show that human travelling behaviour can be described mathematically on many spatiotemporal scales by a two-parameter continuous-time random walk model to a surprising accuracy, and conclude that human travel on geographical scales is an ambivalent and effectively superdiffusive process.

Quantitative aspects of dispersal in ecology are based on the dispersal curve, which quantifies the relative frequency of travel distances of individuals as a function of geographical distance¹¹. A large class of dispersal curves (for example, exponential, gaussian, stretched exponential) permits the identification of a typical length scale by the variance of the displacement length or equivalent quantities. When interpreted as the probability $P(r)$ of finding a displacement of length r in a short time δt , the existence of a typical length scale often justifies the description of dispersal in terms of diffusion equations on large spatiotemporal scales. If, however, $P(r)$ lacks a typical length scale, that is $P(r) \sim r^{-(1+\beta)}$ with $\beta < 2$, the diffusion approximation fails. In physics, random processes with such a single-step distribution are known as Lévy flights^{12–16} (see Supplementary Information). Recently, the related notion of long-distance-dispersal (LDD) has been established in dispersal ecology¹⁷, taking into account the observation that dispersal curves of a number of species show power-law tails owing to long-range movements^{18–21}. (In ecological literature, the term ‘dispersal’ is commonly used in the context of the spatial displacement of individuals of a species between their geographical origin of birth and the location of their first breeding place. Here we use the term dispersal to refer to geographical displacements that occur on much shorter timescales, that is, due to travel by various means of transportation.)

Nowadays, humans travel on many spatial scales, ranging from a few to thousands of kilometres over short periods of time. The direct

quantitative assessment of human movements, however, is difficult, and a statistically reliable estimate of human dispersal comprising all spatial scales does not exist. The central aim of this work is to use data collected at online bill-tracking websites (which monitor the worldwide dispersal of large numbers of individual bank notes) to infer the statistical properties of human dispersal with very high spatiotemporal precision. Our analysis of human movement is based on the trajectories of 464,670 dollar bills obtained from the bill-tracking system www.wheresgeorge.com. We analysed the dispersal of bank notes in the United States, excluding Alaska and Hawaii. The core data consists of 1,033,095 reports to the bill-tracking website. From these reports we calculated the geographical displacements $r = |\mathbf{x}_2 - \mathbf{x}_1|$ between a first ($\mathbf{x}_1$) and secondary ($\mathbf{x}_2$) report location of a bank note and the elapsed time T between successive reports.

In order to illustrate qualitative features of bank note trajectories, Fig. 1b depicts short-time trajectories ($T < 14$ days) originating from three major US cities: Seattle, New York and Jacksonville. After their initial entry into the tracking system, most bank notes are next reported in the vicinity of the initial entry location, that is $|\mathbf{x}_2 - \mathbf{x}_1| \leq 10$ km (Seattle, 52.7%; New York, 57.7%; Jacksonville, 71.4%). However, a small but considerable fraction is reported beyond a distance of 800 km (Seattle, 7.8%; New York, 7.4%; Jacksonville, 2.9%).

From a total of 20,540 short-time trajectories originating across the United States, we measured the probability $P(r)$ of traversing a distance r in a time interval δT of 1–4 days (Fig. 1c). A total of 14,730 (that is, a fraction $Q = 0.71$) secondary reports occurred outside a short range radius $L_{\min} = 10$ km. Between $L_{\min}$ and the approximate average East–West extension of the United States, $L_{\max} \approx 3,200$ km, the kernel shows power-law behaviour $P(r) \sim r^{-(1+\beta)}$ with an exponent $\beta = 0.59 \pm 0.02$. For $r < L_{\min}$, $P(r)$ increases linearly with r , which implies that displacements are distributed uniformly inside the disk $|\mathbf{x}_2 - \mathbf{x}_1| \leq L_{\min}$. We measured $P(r)$ for three classes of initial entry locations: highly populated metropolitan areas (191 sites, local population $N_{\text{loc}} > 120,000$), cities of intermediate size (1,544 sites, local population $120,000 > N_{\text{loc}} > 22,000$) and small towns (23,640 sites, local population $N_{\text{loc}} < 22,000$), comprising 35.7%, 29.1% and 25.2% of the entire population of the United States, respectively. The inset in Fig. 1c shows $P(r)$ for these classes. Despite systematic deviations for short distances, all distributions show an algebraic tail with the same exponent $\beta \approx 0.6$, which confirms that the observed power-law is an intrinsic and universal property of dispersal.

However, the situation is more complex. If we assume that the dispersal of bank notes can be described by a Lévy flight with a short-time probability distribution $P(r)$, we can estimate the time T_{eq} for an initially localized ensemble of bank notes to reach the stationary distribution²² (maps in Fig. 1a), obtaining a value of $T_{\text{eq}} \approx 68$ days (see Supplementary Information). Thus, after 2–3 months, bank notes should have reached an equilibrium distribution. Surprisingly, the long-time dispersal data does not reflect a relaxation within this

¹Max-Planck Institute for Dynamics and Self-Organisation, Bunsenstr. 10, 37073 Göttingen, Germany. ²Department of Physics, University of Göttingen, 37073 Göttingen, Germany. ³Kavli Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA. ⁴Bernstein Center for Computational Neuroscience, 37073 Göttingen, Germany.

time. Figure 1b shows secondary reports of bank notes with initial entry at Omaha that have dispersed for times $T > 100$ days (with an average time $\langle T \rangle = 289$ days). Only 23.6% of the bank notes travelled farther than 800 km, whereas 57.3% travelled an intermediate distance $50 < r < 800$ km, and a relatively large fraction of 19.1% remained within a radius of 50 km, even after an average time of nearly one year. From the computed value $T_{eq} \approx 68$ days, a much higher fraction of bills is expected to reach the metropolitan areas of the West coast and the New England states after this time. This is sufficient evidence that the simple Lévy flight picture for dispersal is incomplete. What causes this attenuation of dispersal?

Two alternative explanations might account for this effect. The slowing down might be caused by strong spatial inhomogeneities of the system. People might be less likely to leave large cities than for example, suburban areas. Alternatively, long periods of rest might be an intrinsic temporal property of dispersal. In as much as an algebraic tail in spatial displacements yields superdiffusive behaviour, a tail in the probability density $\phi(t)$ for times t between successive spatial displacements of an ordinary random walk can lead to subdiffusion¹⁵

(see Supplementary Information). Here, the ambivalence between scale-free spatial displacements and scale-free periods of rest can be responsible for the observed attenuation of superdiffusion.

In order to address this issue we investigated the relative proportion $P_0^i(t)$ of bank notes which are reported in a small (20 km) radius of the initial entry location i as a function of time (Fig. 1d). The quantity $P_0^i(t)$ estimates the probability of a bank note being reported at the initial location at time t . We computed $P_0^i(t)$ for metropolitan areas, cities of intermediate size and small towns: for all classes we found the asymptotic behaviour $P_0(t) \sim At^{-\eta}$, with the same exponent $\eta = 0.6 \pm 0.03$, which indicates that waiting time and dispersal characteristics are universal. Notice that for a pure Lévy flight with index β in two dimensions, $P_0(t)$ scales with time as $t^{-2/\beta}$ (dashed red line)¹⁵. For $\beta \approx 0.6$ (as suggested by Fig. 1c) this implies $\eta \approx 3.33$. This is a fivefold steeper decrease than observed, which clearly shows that dispersal cannot be described by a pure Lévy flight. The measured decay is even slower than the decay expected from ordinary two-dimensional diffusion ($\eta = 1$, dashed black line). Therefore, we conclude that the slow decay in $P_0(t)$ reflects the effect

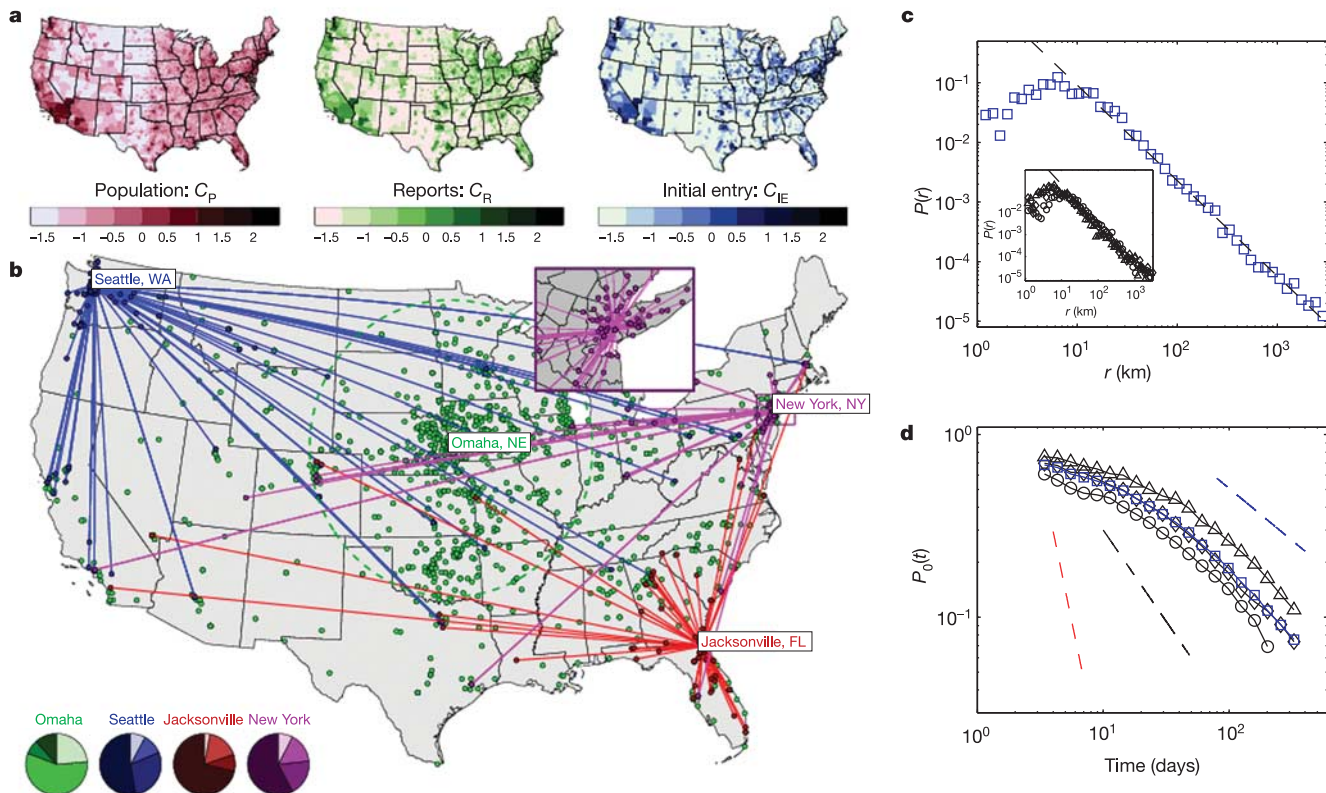


Figure 1 | Dispersal of bank notes and humans on geographical scales. **a**, Relative logarithmic densities of population ($c_P = \log \rho_P / \langle \rho_P \rangle$), report ($c_R = \log \rho_R / \langle \rho_R \rangle$) and initial entry ($c_{IE} = \log \rho_{IE} / \langle \rho_{IE} \rangle$) as functions of geographical coordinates. Colour-code shows densities relative to the nationwide averages (3,109 counties) of $\langle \rho_P \rangle = 95.15$, $\langle \rho_R \rangle = 0.34$ and $\langle \rho_{IE} \rangle = 0.15$ individuals, reports and initial entries per km², respectively. **b**, Trajectories of bank notes originating from four different places. City names indicate initial location, symbols secondary report locations. Lines represent short-time trajectories with travelling time $T < 14$ days. Lines are omitted for the long-time trajectories (initial entry in Omaha) with $T > 100$ days. The inset depicts a close-up view of the New York area. Pie charts indicate the relative number of secondary reports coarsely sorted by distance. The fractions of secondary reports that occurred at the initial entry location (dark), at short ($0 < r < 50$ km), intermediate ($50 < r < 800$ km) and long ($r > 800$ km) distances are ordered by increasing brightness of hue. The total number of initial entries are $N = 2,055$ (Omaha), $N = 524$ (Seattle), $N = 231$ (New York), $N = 381$ (Jacksonville). **c**, The short-time

dispersal kernel. The measured probability density function $P(r)$ of traversing a distance r in less than $T = 4$ days is depicted in blue symbols. It is computed from an ensemble of 20,540 short-time displacements. The dashed black line indicates a power law $P(r) \sim r^{-(1+\beta)}$ with an exponent of $\beta = 0.59$. The inset shows $P(r)$ for three classes of initial entry locations (black triangles for metropolitan areas, diamonds for cities of intermediate size, circles for small towns). Their decay is consistent with the measured exponent $\beta = 0.59$ (dashed line). **d**, The relative proportion $P_0(t)$ of secondary reports within a short radius ($r_0 = 20$ km) of the initial entry location as a function of time. Blue squares show $P_0(t)$ averaged over 25,375 initial entry locations. Black triangles, diamonds, and circles show $P_0(t)$ for the same classes as **c**. All curves decrease asymptotically as $t^{-\eta}$ with an exponent $\eta = 0.60 \pm 0.03$ indicated by the blue dashed line. Ordinary diffusion in two dimensions predicts an exponent $\eta = 1.0$ (black dashed line). Lévy flight dispersal with an exponent $\beta = 0.6$ as suggested by **b** predicts an even steeper decrease, $\eta = 3.33$ (red dashed line).

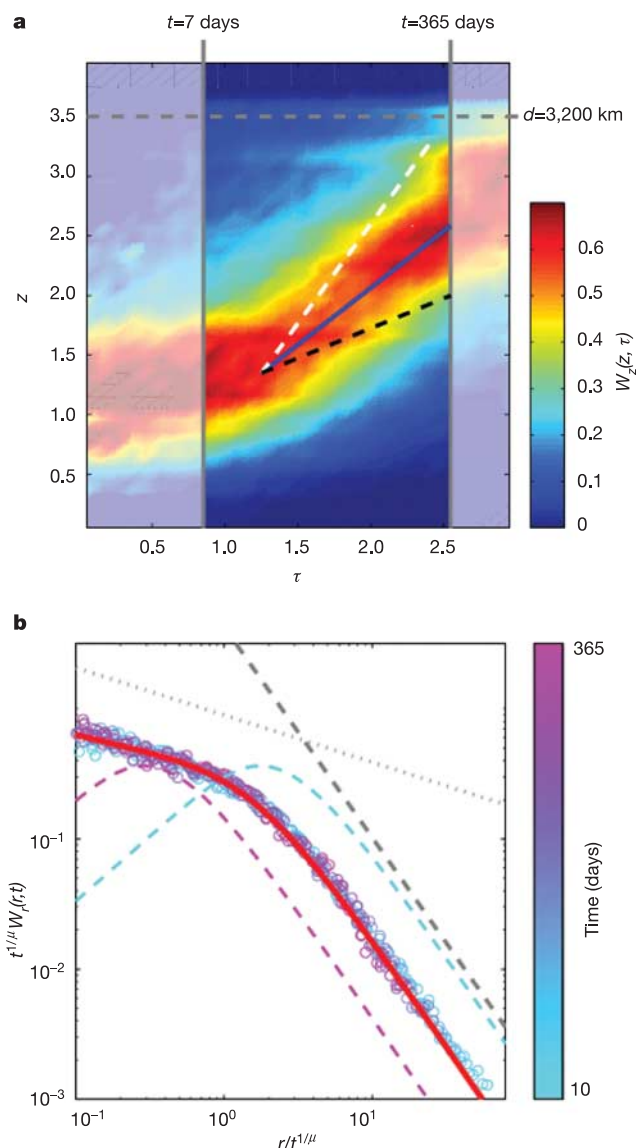


Figure 2 | Spatiotemporal scaling of bank note dispersal. **a**, The probability density $W_z(z, \tau)$ of having travelled a logarithmic distance $z = \log_{10} r$ at logarithmic time $\tau = \log_{10} t$. The middle segment indicates the scaling regime between one week and one year. The superimposed red line represents the scaling behaviour $r(t) \sim t^{1/\mu}$ with exponent $\mu = 1.05 \pm 0.05$. It is compared to the diffusive scaling (black dashed line) and the scaling of a pure Lévy process with exponent $\beta = 0.6$ (white dashed line). The upper dashed grey line shows the approximate linear extent $L_{\max} = 3,200$ km of the United States. **b**, The measured radial probability density $W_r(r, t)$ and theoretical scaling function $L_{\alpha, \beta}(r/t^{1/\mu})$ (equation (2)). In order to extract the quality of scaling, the function $t^{1/\mu} W_r(r, t)$ is shown for various but fixed values of t from 10–365 days as a function of the rescaled distance $r/t^{1/\mu}$, where the exponent μ was set to the value determined in **a**. As the measured (circles) curves collapse on a single curve, the process shows universal scaling. The scaling curve represents the limiting density of the process. The asymptotic behaviour for small (grey dotted line) and large (grey dashed line) arguments $y = r/t^{1/\mu}$ is given by $y^{-(1-\xi_1)}$ and $y^{-(1+\xi_2)}$, respectively, with estimated exponents $\xi_1 = 0.63 \pm 0.04$ and $\xi_2 = 0.62 \pm 0.02$. According to our model, these exponents must fulfill $\xi_1 = \xi_2 = \beta$, where β is the exponent of the asymptotic short-time dispersal kernel (Fig. 1c), that is $\beta \approx 0.6$. The superimposed red line represents $t^{1/\mu} W_r(r, t)$ predicted by our theory, with spatial and temporal exponents $\alpha = 0.6$ and $\beta = 0.6$, respectively. The coloured dashed lines represent $t^{1/\mu} W_r(r, t)$ for a pure Lévy flight with $\beta = 0.6$ at times $t = 10$ and $t = 365$ days. The curves do not collapse because the pure Lévy flight shows the wrong spatiotemporal scaling. Furthermore, the limiting curves strongly deviate from the data for small arguments.

of an algebraic tail in the distribution of rests $\phi(t)$ between displacements. Indeed, if $\phi(t) \sim t^{-(1+\alpha)}$ with $\alpha < 1$, then $\eta = \alpha$ and consequently $\alpha = 0.60 \pm 0.03$. This suggests that an algebraic tail in the distribution of rests $\phi(t)$ is responsible for slowing down the superdiffusive dispersal advanced by the short time dispersal kernel in Fig. 1c.

In order to model the antagonistic interplay between scale-free displacements and waiting times, we use the framework of continuous-time random walks (CTRW) introduced by Montroll and Weiss²³. A CTRW consists of a succession of random displacements $\delta \mathbf{x}_n$ and random waiting times δt_n , each of which is drawn from a corresponding probability density function $P(\delta \mathbf{x}_n)$ and $\phi(\delta t)$. After N iterations, the position of the walker and the elapsed time are given by $\mathbf{x}_N = \sum_n \delta \mathbf{x}_n$ and $t_N = \sum_n \delta t_n$. The quantity of interest is the position $\mathbf{x}(t)$ after time t and the associated probability density $W(\mathbf{x}, t)$ that can be computed within CTRW theory. For displacements with finite variance σ^2 and waiting times with finite mean τ , such a CTRW yields ordinary diffusion asymptotically, that is $\partial_t W(\mathbf{x}, t) = D \partial_x^2 W(\mathbf{x}, t)$ with a diffusion coefficient $D = \sigma^2/\tau$.

In contrast, we assume here that both $P(\delta \mathbf{x}_n)$ and $\phi(\delta t)$ show algebraic tails, that is $P(\delta \mathbf{x}_n) \sim |\delta \mathbf{x}_n|^{-(1+\beta)}$ and $\phi(\delta t) \sim |\delta t|^{-(1+\alpha)}$, for which σ^2 and τ are infinite. In this case we can derive a bifractional diffusion equation for the dynamics of $W(\mathbf{x}, t)$:

$$\partial_t^\alpha W(\mathbf{x}, t) = D_{\alpha, \beta} \partial_{|\mathbf{x}|}^\beta W(\mathbf{x}, t) \quad (1)$$

In this equation, the symbols ∂_t^α and $\partial_{|\mathbf{x}|}^\beta$ denote fractional derivatives that are non-local and depend on the tail exponents α and β . The constant $D_{\alpha, \beta}$ is a generalized diffusion coefficient (see Supplementary Information). Equation (1) represents the core dynamical equation of our model. Using methods of fractional calculus we can solve this equation and obtain the probability $W_r(r, t)$ of having traversed a distance r at time t :

$$W_r(r, t) = t^{-\alpha/\beta} L_{\alpha, \beta}(r/t^{\alpha/\beta}) \quad (2)$$

where $L_{\alpha, \beta}$ is a universal scaling function that represents the characteristics of the process. Equation (2) implies that the typical distance travelled scales according to $r(t) \sim t^{1/\mu}$, where $\mu = \beta/\alpha$. Thus, depending on the ratio of spatial and temporal exponents, the random walk can be effectively either superdiffusive ($\beta < 2\alpha$), subdiffusive ($\beta > 2\alpha$), or quasidiffusive ($\beta = 2\alpha$) (see Supplementary Information). For the exponents observed in the dispersal data ($\beta = 0.59 \pm 0.02$ and $\alpha = 0.60 \pm 0.03$) the theory predicts a temporal scaling exponent in the vicinity of unity, $\mu = 0.98 \pm 0.08$. Therefore, dispersal remains superdiffusive despite long periods of rest.

The validity of our model can be tested by estimating $W_r(r, t)$ from the entire data set of a little over half a million displacements and elapse times. The scaling property is best extracted from the data by a transformation to logarithmic coordinates $z = \log_{10} r$, $\tau = \log_{10} t$ and the associated probability density $W_z(z, \tau)$. If the original process scales according to $r(t) \sim t^{1/\mu}$, the density $W_z(z, \tau)$ is a function of $z - \tau/\mu$ only. Figure 2a shows that scaling occurs in a time window of approximately seven days to one year. From the slope (blue line), we obtain a scaling exponent $\mu = 1.05 \pm 0.02$, which agrees well with our model.

Finally, we investigated the degree to which bank note dispersal shows a scaling density as predicted by our model (that is, the relation outlined in equation (2)). Figure 2b shows $t^{1/\mu} W_r(r, t)$ extracted from data versus the ratio $y = r/t^{1/\mu}$. The exponent $\mu = 1.05$ was set to the value obtained in Fig. 2a. The collapse of the data on a single curve indicates that in the chosen time interval of 10–365 days, bank note dispersal shows a universal scaling function. The asymptotic behaviour of the empirical curve is given by $y^{-(1-\xi_1)}$ and $y^{-(1+\xi_2)}$ for small and large arguments, respectively. Both exponents fulfil $\xi_1 \approx \xi_2 \approx 0.6$. We compared the empirical curve with the theoretical prediction of our model. By series expansions, we can compute the asymptotics of the limiting function $L_{\alpha, \beta}(y)$ in equation (2),

giving $y^{-(1-\beta)}$ and $y^{-(1+\beta)}$ for small and large y , respectively. Consequently, as $\beta \approx 0.6$ (Fig. 1c), the theory agrees well with the observed exponents. For the entire range of y we computed $L_{\alpha,\beta}(y)$ by numeric integration for $\alpha = \beta = 0.6$, and superimposed the theoretical curve on the empirical one. The agreement is very satisfactory. In summary, our analysis gives solid evidence that the dispersal of bank notes can be accounted for by our model.

The question remains how the dispersal characteristics of bank notes carry over to the travelling behaviour of humans. In this context, we can conclude that the power law with exponent $\beta = 0.6$ of the short-time dispersal kernel for bank notes reflects the human dispersal kernel, because the exponent remains unchanged for short time intervals of $T = 2, 4, 7$ and 14 days. The issue of long waiting times is more subtle. One might speculate that the observed algebraic tail in waiting times of bank notes is a property of bank note dispersal alone. Long waiting times might be caused by bank notes that exit the money-tracking system for a long time, for instance in banks. However, if this were the case the inter-report time statistics would have an algebraic tail as well. Analysing the inter-report time distribution, we found an exponential decay, suggesting that bank notes are passed from person to person at a constant rate. If we assume that humans exit small areas at a constant rate that is equivalent to exponentially distributed waiting times, and that bank notes pass from person to person at a constant rate, the distribution of bank note waiting times would also be exponential, in contrast to the observed power law. To our minds, this reasoning permits no other conclusion than a lack of scale in human waiting-time statistics. We obtained further support for our results from a comparison with two independent human travel data sets: long-distance travel on the United States aviation network⁸ (flight schedules and airport information, www.oag.com; International Air Transport Association, www.iata.org) and the latest survey on long-distance travel conducted by the United States Bureau of Transportation Statistics (www.bts.gov) (see Supplementary Information). Both agree well with our findings and support our conclusions.

On the basis of our analysis, we conclude that the dispersal of bank notes and human travel behaviour can be described by a continuous-time random-walk process that incorporates scale-free jumps as well as long waiting times between displacements. To our knowledge, this is the first empirical evidence for such an ambivalent process in nature. We believe that these results can serve as a starting point for the development of a new class of models for the spread of human infectious diseases, because universal features of human travel can now be accounted for in a quantitative way.

Received 13 July; accepted 3 October 2005.

1. Bullock, J. M., Kenward, R. E. & Hails, R. S. (eds) *Dispersal Ecology* (Blackwell, Malden, Massachusetts, 2002).
2. Murray, J. D. *Mathematical Biology* (Springer-Verlag, New York, 1993).

3. Clobert, J., Danchin, E., Dhondt, A. A. & Nichols, J. D. (eds) *Dispersal* (Oxford Univ. Press, Oxford, 2001).
4. Nicholson, K. & Webster, R. G. *Textbook of Influenza* (Blackwell, Malden, Massachusetts, 1998).
5. Grenfell, B. T., Bjornstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001).
6. Keeling, M. J. et al. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813–817 (2001).
7. Hudson, P. J., Rizzoli, A., Grenfell, B. T. & Heesterbeek, H. (eds) *The Ecology of Wildlife Diseases* (Oxford Univ. Press, Oxford, 2002).
8. Hufnagel, L., Brockmann, D. & Geisel, T. Forecast and control of epidemics in a globalized world. *Proc. Natl Acad. Sci. USA* **101**, 15124–15129 (2004).
9. Grassly, N. C., Fraser, C. & Garnett, G. P. Host immunity and synchronized epidemics of syphilis across the United States. *Nature* **433**, 417–421 (2005).
10. Webby, R. J. & Webster, R. G. Are we ready for pandemic influenza? *Science* **302**, 1519–1522 (2003).
11. Kot, M., Lewis, M. A. & van den Driessche, P. Dispersal data and the spread of invading organisms. *Ecology* **77**, 2027–2042 (1996).
12. Shlesinger, M. F., Zaslavsky, G. M. & Frisch, U. (eds) *Lévy Flights and Related Topics in Physics* (Springer Verlag, Berlin, 1995).
13. Klafter, J., Shlesinger, M. F. & Zumofen, G. Beyond Brownian motion. *Phys. Today* **49**, 33–39 (1996).
14. Brockmann, D. & Geisel, T. Lévy flights in inhomogeneous media. *Phys. Rev. Lett.* **90**, 170601 (2003).
15. Metzler, R. & Klafter, J. The random walks guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **339**, 1–77 (2000).
16. Shlesinger, M. F., Klafter, J. & Wong, Y. M. Random-walks with infinite spatial and temporal moments. *J. Stat. Phys.* **27**, 499–512 (1982).
17. Nathan, R. The challenges of studying dispersal. *Trends Ecol. Evol.* **16**, 481–483 (2001).
18. Viswanathan, G. M. et al. Lévy flight search patterns of wandering albatrosses. *Nature* **381**, 413–415 (1996).
19. Ramos-Fernández, G., Mateos, J. L., Miramontes, O. & Cocho, G. Lévy walk patterns in the foraging movements of spider monkeys. *Behav. Ecol. Sociobiol.* **55**, 223–230 (2004).
20. Levin, S. A., Muller-Landau, H. C., Nathan, R. & Chave, J. The ecology and evolution of seed dispersal: A theoretical perspective. *Annu. Rev. Ecol. Syst.* **34**, 575–604 (2003).
21. Nathan, R. et al. Mechanisms of long-distance dispersal of seeds by wind. *Nature* **418**, 409–413 (2002).
22. Gardiner, C. W. *Handbook of Stochastic Methods* (Springer Verlag, Berlin, 1985).
23. Montroll, E. W. & Weiss, G. H. Random walks on lattices. *J. Math. Phys.* **6**, 167–181 (1965).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank the initiators of the bill tracking system (www.wheresgeorge.com). We thank cabinetmaker D. Derryberry for discussions and for drawing our attention to the [wheresgeorge](http://wheresgeorge.com) website, and B. Shraiman, D. Cohen and W. Noyes for critical comments on the manuscript.

Author Contributions The project idea was conceived by D.B. and L.H., data pre-processing was done by L.H., data analysis by D.B. and L.H., the theory and model was constructed by D.B., and the manuscript was written by D.B., L.H. and T.G.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.B. (brockmann@ds.mpg.de).

LETTERS

Empathic neural responses are modulated by the perceived fairness of others

Tania Singer^{1,2}, Ben Seymour¹, John P. O'Doherty³, Klaas E. Stephan¹, Raymond J. Dolan¹ & Chris D. Frith¹

The neural processes underlying empathy are a subject of intense interest within the social neurosciences^{1–3}. However, very little is known about how brain empathic responses are modulated by the affective link between individuals. We show here that empathic responses are modulated by learned preferences, a result consistent with economic models of social preferences^{4–7}. We engaged male and female volunteers in an economic game, in which two confederates played fairly or unfairly, and then measured brain activity with functional magnetic resonance imaging while these same volunteers observed the confederates receiving pain. Both sexes exhibited empathy-related activation in pain-related brain areas (fronto-insular and anterior cingulate cortices) towards fair players. However, these empathy-related responses were significantly reduced in males when observing an unfair person receiving pain. This effect was accompanied by increased activation in reward-related areas, correlated with an expressed desire for revenge. We conclude that in men (at least) empathic responses are shaped by valuation of other people's social behaviour, such that they empathize with fair opponents while favouring the physical punishment of unfair opponents, a finding that echoes recent evidence for altruistic punishment.

Empathy enables us to share the emotion, pain and sensation of others. The perception–action model of empathy states that the observation or imagination of another person in a particular emotional state automatically activates a representation of that state in the observer¹. Recent imaging studies provide evidence for common activation elicited when experiencing disgust⁸, touch⁹ or pain^{10–12} in oneself, and when perceiving the same feelings in others. For example, studies on empathy for pain have found that the activation observed in anterior insula/fronto-insular cortex (AI/FI) and the anterior cingulate cortex (ACC) to pain in oneself is also seen when observing pain in someone else^{10,12}. These results suggest that our ability to empathize relies on neuronal systems that underpin our own bodily and emotional states^{1,2,10,13}. However, it is unclear whether or how these responses are modulated by the social relations between individuals.

To address this question, we measured brain responses when individuals empathized with the pain of someone they liked or disliked. We used an economic game model to induce liking or disliking of two confederate actors, previously unknown to our experimental subjects. The confederates played fair or unfair strategies in a sequential Prisoner's Dilemma game (PDG) with the subjects. As illustrated in Fig. 1a, in this game the subject was always 'first mover' and could either trust the other player by sending his/her money to him/her or mistrust him/her by keeping it. The confederates were always 'second mover' and could choose between a fair or an unfair response by returning high or low amounts of money (Supplementary Information). On the basis of previous findings¹⁴,

we expected that subjects would come to like fair players and dislike unfair players.

As Fig. 1b illustrates, subjects perceived the confederates as being fair and unfair according to their game-playing strategy. Post-scan behavioural ratings confirmed that both male and female subjects rated the fair player as being significantly more fair, more agreeable, more likeable and more attractive than the unfair player (range of scale +2 to –2; $P < 0.001$; Supplementary Table S1).

In the second part of the experiment we used functional magnetic resonance imaging (fMRI) to investigate whether the liking or disliking acquired during the preceding game modulated empathic responses for pain. One actor sat on each side of the scanner, enabling the scanned volunteer to observe the hands of the fair and unfair player as well as his/her own hand. Painful stimulation was applied through electrodes to the hands of all three participants. As shown in Fig. 1c, cues (coloured arrows) were presented in random order indicating whether she/he (self condition), the fair player (fair condition) or the unfair player (unfair condition) would get low stimulation (no pain condition) or high stimulation (pain condition).

We predicted that pain-related empathic responses in AI/FI and ACC would be elicited when observing a fair person in pain but that this activity would be reduced or absent when observing pain in a person who had previously played unfairly. In view of recent economic models of social preferences^{4–7} and altruistic punishment^{15–17} we further expected that the reduction of empathy for an unfair person might be accompanied by an increase in activity in brain areas known to have a key function in reward processing, such as ventral striatum/nucleus accumbens and orbito-frontal cortex^{18–20}. It has been shown that people reward others for cooperative behaviour but punish violations of social fairness even at a personal cost^{15,16}, an effect likely to be mediated by neural mechanisms that provide intrinsic motivation (that is, satisfaction) from punishing violators²¹.

The comparison of brain activity associated with painful and non-painful trials in the 'self' condition for men and women revealed an expected increase in the 'pain network', including activity in AI and ACC (Supplementary Tables 2 and 3). Moreover, we observed pain-related empathic responses in both genders in bilateral AI extending into FI and brainstem when seeing an unfamiliar but likeable person in pain (Fig. 2a, b). The activation in ACC was significant in women ($[9, 18, 27]$; $P < 0.001$ uncorrected, $P < 0.05$ whole-brain corrected) but was borderline in men ($[-9, 39, 27]$; $P = 0.001$ uncorrected; Supplementary Tables 4 and 5; coordinates refer to the peak of activations in MNI (Montreal Neurological Institute) space). Furthermore, we extended previous findings¹⁰ by showing that men, as well as women, who scored higher on standard empathy scales²² had higher empathy-related brain activity in ACC and AI/FI (Supplementary Fig. 2).

¹Wellcome Department of Imaging Neuroscience and ²Institute of Cognitive Neuroscience, University College of London, London WC1N 3AR, UK. ³Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125, USA.

To investigate whether empathic responses are modulated as a function of the perceived fairness of others we identified, for men and women, peak voxels of activation in bilateral FI observed when pain was applied to self and to fair players from the analysis described above (see Fig. 2a, b). Figure 2c, d illustrates the average activation (parameter estimates) for painful–non-painful stimulation in these voxels when subjects observed either fair or unfair players in receipt of pain. This analysis revealed that less empathic activity was elicited by the knowledge that an unfair player was in pain. However, there was also a marked difference between the sexes. In women this reduction in activity was very small, whereas in men the knowledge

that an unfair player was in receipt of pain elicited no increase in empathic activity in FI. And indeed, formal analysis revealed no significant difference for women when comparing painful trials for fair versus unfair players in empathy-related pain regions. However, men showed significantly enhanced activation in bilateral FI when observing fair compared with unfair players in pain (Supplementary Tables 8 and 11). Consistent with this finding, supplementary analysis showed that women but not men displayed significant activation in bilateral AI/FI and ACC in all three conditions (Supplementary Tables 7 and 10).

We next sought evidence of increases in brain activity in regions associated with reward processing (ventral striatum/nucleus accumbens and orbito-frontal cortex) when observing an unfair player receiving painful stimulation. As shown in Fig. 3a, we observed increased activation in left ventral striatum/nucleus accumbens for men, but not for women, in the comparison of painful trials in the unfair and the fair condition (pain in unfair–pain in fair) as well as in the interaction ((pain–no pain) unfair–(pain–no pain) fair). This latter comparison also revealed increased activation in left orbito-frontal cortex (Supplementary Table 11). As illustrated in Fig. 3b, the direct gender comparison for painful trials in the unfair and the fair condition ((pain in unfair–pain in fair) men–(pain in unfair–pain in fair) women) revealed that men showed significantly higher activation than women in left nucleus accumbens (in the same peak as identified in the former analysis).

To explore the role of nucleus accumbens and orbito-frontal cortex further, we assessed whether individual differences in an expressed desire for revenge covaried with brain activity in these regions. We derived a ‘revenge’ composite score from post-experiment questionnaires measuring a subjectively expressed desire for revenge (see Methods and Supplementary Information). Figure 3c shows that men expressed a stronger desire for revenge than women ($t(30) = 2.40$, $P < 0.05$; Supplementary Fig. 3). As illustrated in Fig. 3d, regression analysis confirmed that men, but not women, who expressed a stronger desire for revenge showed greater activation in nucleus accumbens when they perceived an unfair player receiving painful stimuli than when they perceived a fair player in pain (Supplementary Fig. 4).

Our data provide neurobiological evidence on how fairness in social interactions shapes the nature of the affective link between people. Our findings indicate that cooperation nourishes this link, but selfish behaviour that is detrimental to others effectively compromises this link (at least with males), such that empathic responses in the brain are diminished or abolished.

These findings complement those of a previous imaging study that reported enhanced activation in dorsal striatum (caudate nucleus) when individuals punished defectors (by delivering ‘punishment points’) in a sequential Prisoner’s Dilemma game²¹. In the present study we observed activation correlated with revenge in ventral striatum. This difference in evoked activity in dorsal and ventral regions of the striatum is likely to reflect the different nature of the tasks used. In the previous study²¹ subjects were required to select an action to administer punishment, whereas in the present study subjects passively observed a cue indicating that a defector was receiving pain. These findings are consistent with the different functions associated with distinct regions of the striatum: afferent projections to dorsal striatum are thought to be crucial for learning correct actions so as to maximize reward, whereas projections to ventral striatum, including nucleus accumbens, have a key function in reward prediction and pavlovian learning^{18–20,23,24}. The findings of enhanced activation in ventral striatum to a signal indicating that a defector is receiving pain are in agreement with the hypothesis that humans derive satisfaction simply from seeing justice administered^{15,21}, even if the instrument of punishment is out of their control.

Our results suggest a neural foundation for theories of social preferences. These theories^{4,7} suggest that people value the gains of others positively if they are perceived to behave fairly, but value

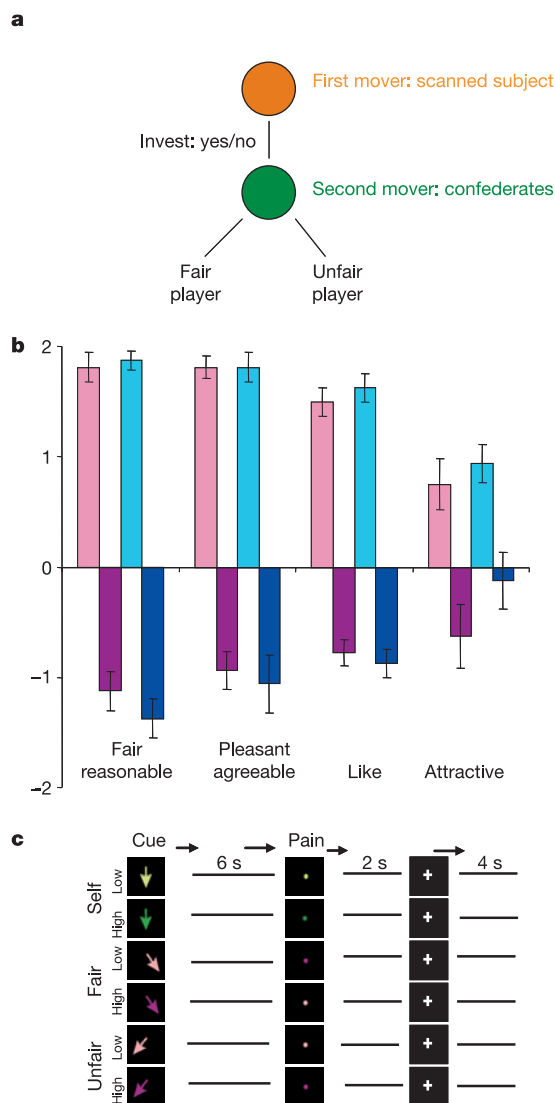


Figure 1 | Experimental models I and II and behavioural ratings. **a**, Game tree of a sequential Prisoner's Dilemma game used to induce liking and disliking of opponent players. First movers (subjects) chose between keeping and sending money to the second mover. One of the confederates sent fair (large) amounts of money back, the other unfair (small) amounts. **b**, Post-scan ratings showing that both sexes rated the fair confederates as being more reasonable/fair, their personality as being more agreeable/pleasant, as being more likeable and more attractive than unfair players (scale ranging from -2 to +2). Pink bars, women subjects rating the fair player; purple bars, women subjects rating the unfair player; cyan bars, men subjects rating the fair player; blue bars, men subjects rating the unfair player. Error bars refer to s.e.m. **c**, Experimental design of the 'empathy for pain model'. Arrows of different colours indicate painful or non-painful stimulation applied to the scanned subject (self), the fair or the unfair player.

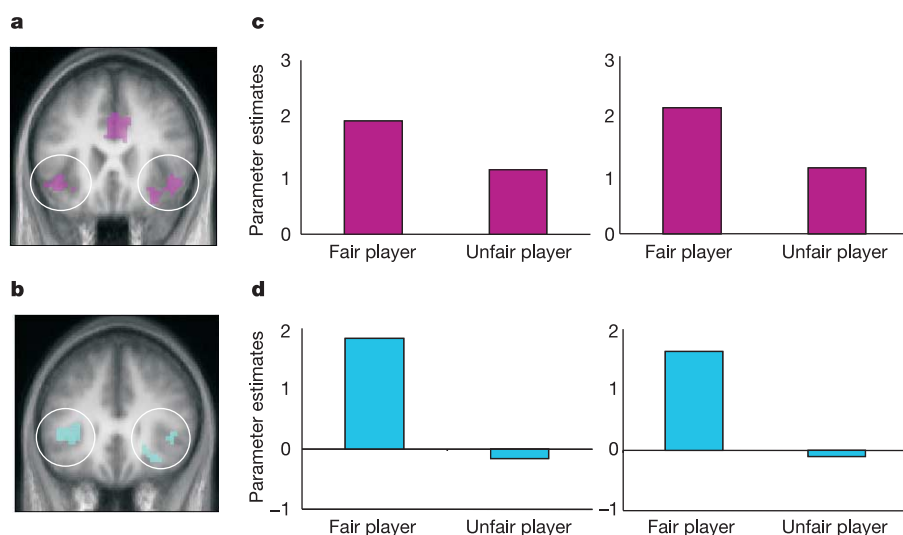


Figure 2 | Pain-sensitive activation networks to the sight of fair and unfair players in pain. **a, b,** Conjunction analysis between the contrasts pain–no pain in the context of self and the fair condition at $P < 0.001$ for women (pink, **a**) and men (blue, **b**). Increased pain-related activation (asterisk indicates whole-brain corrected) for women in ACC* [9, 18, 27], left FI* [–42, 15, –3], right FI* [30, 18, –18], left SII* [–60, –30, 18], right SII*

[63, –30, 24] and brainstem* [3, –18, –18]; for men in left FI* [–33, 33, 3] right FI [42, 33, 3] and brainstem [3, –33, –30]. **c, d,** Average activation (parameter estimates) in peak voxels of left and right FI (left and right panels, respectively) for the painful–non painful trials in fair and unfair conditions for women (**c**) and men (**d**).

others' gains negatively if they behave unfairly. This pattern of preferences implies that people like cooperating with fair opponents but also like punishing unfair opponents. Our corresponding neurobiological observations were more prominent in men, although further experiments are needed to confirm the gender specificity of the effect. It is possible that our experimental design favoured men because the modality of punishment was related to physical threat, as opposed to psychological or financial threat. Alternatively, these findings could indicate a predominant role for males in the maintenance of justice and punishment of norm violation in human societies.

METHODS

A full description of experimental design and methods is provided in Supplementary Information, and an overview of the design is provided in Supplementary Fig. 1.

Subjects and confederates. Subjects (16 men and 16 women) gave informed consent and the study was approved by the Joint Ethics Committee of the National Hospital for Neurology and Neurosurgery (University College London Hospitals NHS Foundation Trust) and Institute of Neurology (University College London). We employed four professional actors (two female and two male) as confederates.

Induction of liking and disliking. Subjects played a sequential iterated Prisoner's Dilemma game, in which a first player can trust a second player by sending 10 starting points (subsequently converted to money) to the other player knowing that each point sent will be tripled²⁵. The second player (confederate) then reciprocates by sending an amount between 0 and 10 points back, which is also tripled. In each game, two confederates, sitting in different rooms, were playing fairly or unfairly by reciprocating large or small amounts of money.

Empathy for pain model. We used an electrical pain stimulus applied to the dorsum of the right hand¹⁰. Before the experiment we determined individual current amplitudes for high and low intensity stimulation, for the subject and both confederates (see Supplementary Table 1). The two confederates sat either side of the subject (positioned in the scanner) allowing the subject, with the aid of mirrors, to see all three hands. Each 12-s trial involved presentation of a visual cue, which was followed after 6 s by a small circle (for 2 s) indicating the beginning of electrical stimulation (Fig. 1c). The cue indicated whether the subject (self), the fair player or the unfair player would get low stimulation (no pain) or high stimulation (pain). Stimulation intensity was indicated by the colour intensity of the cue. Each empathy for pain session consisted of ten trials

of each of the six conditions (pain and no-pain in the context of self, fair and unfair) and 20 null events.

Although the two confederates were always of the same sex, all four possible gender combinations between the sex of the subject and the sex of the confederates were used equally often throughout the study. The position (left or right) of the actor (fair or unfair) was counterbalanced across subjects.

Post-scan questionnaires. After scanning, subjects completed a standard empathy scale²² and were asked to rate the intensity of the low and high stimulation, their liking for the two confederates, and their desire for revenge on the two confederates (see Fig. 1b, Supplementary Table 1 and Supplementary Fig. 3).

Image acquisition and analysis. We used a 1.5-T Siemens Sonata MRI scanner to acquire gradient-echo, T_2^* -weighted echo-planar images with blood-oxygenation-level-dependent contrast. An additional T_1 -weighted structural image was acquired for each subject.

After correction for head movements and spatial normalization²⁶, the images were analysed with SPM2 (Wellcome Department of Imaging Neuroscience, London) using an event-related model²⁷. The experiment constituted a $2 \times 3 \times 2$ factorial design with the first factor representing 'intensity of stimulation' (pain versus no pain), the second factor 'addressee' (self, fair and unfair) and the third factor 'gender' (male or female subject).

To create regressors of interest, each condition was modelled by convolving delta functions at each trial onset (presentation of the anticipatory cue) and at each pain onset (presentation of the circle) with a canonical haemodynamic response function. Contrast images were calculated by applying linear contrasts to the parameter estimates for the regressor of each event. The contrast images were then entered into one-sample t -tests, separately for female and male subjects, to instantiate random-effects group analyses^{28,29}.

To assess shared networks of pain-related activation in self, fair and unfair conditions, we performed a conjunction analysis and an additional (more conservative) inclusive masking procedure in which we masked the pain–no-pain contrast in the fair or unfair condition with the pain–no-pain contrast in self.

Finally, we used a regression analysis to explore which brain regions showed, first, a correlation between empathy-related activity (pain–no pain in fair) and individual empathic character traits as assessed by post-scan empathy questionnaires, and second, a correlation between pain-related activity in the unfair condition compared with the fair condition (that is, pain in unfair–pain in fair) and individual tendencies to seek revenge as assessed by subjective rating scales.

On the basis of previous findings^{10–12} we expected two regions to be of particular importance for pain-related empathy: FI and ACC. In the analysis focusing on activity specific to perceiving pain in unfair players in comparison

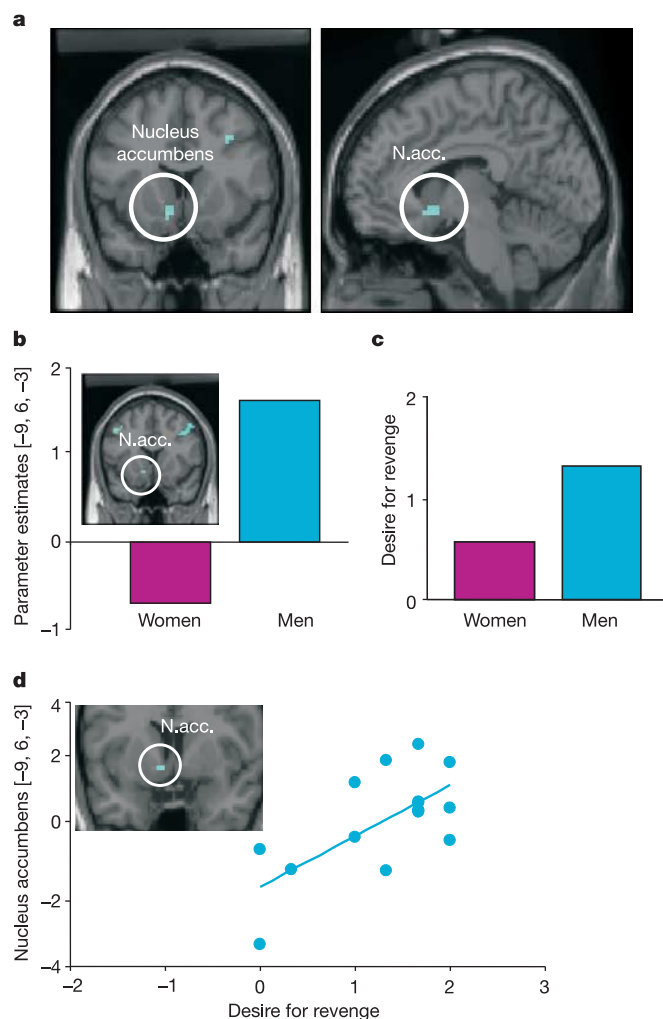


Figure 3 | Gender differences in brain activity in nucleus accumbens specific to the perception of an unfair compared to fair player in pain. **a**, Increased activity ($P < 0.005$) in nucleus accumbens $[-9, 15, -9]$ for painful trials in the unfair–fair condition for men but not for women. **b**, Average activation (parameter estimates) for women (pink) and men (blue) in left nucleus accumbens $[-9, 15, -9]$ when testing for gender differences. **c**, Men (blue) indicate stronger feelings of desire for revenge than women (pink) ($t(30) = 2.40$, $P < 0.05$) measured on a scale from -2 ('not at all') to $+2$ ('very much'). **d**, Correlation ($r = 0.68$, $P < 0.05$) of parameter estimates at peak of nucleus accumbens activation $[-9, 6, -3]$ for the (pain in unfair–pain in fair) contrast in men with expressed desire for revenge in men. There was no correlation for women.

with fair players we extended our regions of interest to include areas known to be involved in reward processing including ventral striatum (nucleus accumbens) and orbito-frontal cortex. We report results at $P < 0.005$ uncorrected for multiple comparisons in the a priori regions of interest.

Received 15 June; accepted 30 September 2005.
Published online 18 January 2006.

- Preston, S. D. & de-Waal, F. B. M. Empathy: Its ultimate and proximate bases. *Behav. Brain Sci.* **25**, 1–72 (2002).
- Gallese, V., Keysers, C. & Rizzolatti, G. A unifying view of the basis of social cognition. *Trends Cogn. Sci.* **8**, 396–403 (2004).

- Decety, J. & Jackson, P. L. The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* **3**, 71–100 (2004).
- Rabin, M. Incorporating fairness into game theory and economics. *Am. Econ. Rev.* **83**, 1281–1302 (1993).
- Levine, D. K. Modeling altruism and spitefulness in experiments. *Rev. Econ. Dynam.* **1**, 593–622 (1998).
- Camerer, C. F. *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton Univ. Press, Princeton, 2003).
- Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868 (1999).
- Wicker, B. *et al.* Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron* **40**, 655–664 (2003).
- Keysers, C. *et al.* A touching sight: SII/PV activation during the observation and experience of touch. *Neuron* **42**, 335–346 (2004).
- Singer, T. *et al.* Empathy for pain involves the affective but not sensory components of pain. *Science* **303**, 1157–1162 (2004).
- Jackson, P. L., Meltzoff, A. N. & Decety, J. How do we perceive the pain of others: A window into the neural processes involved in empathy. *Neuroimage* **3**, 771–779 (2005).
- Morrison, I., Lloyd, D., di Pellegrino, G. & Roberts, N. Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue? *Cogn. Affect. Behav. Neurosci.* **4**, 270–278 (2004).
- Craig, A. D. Human feelings: why are some more aware than others? *Trends Cogn. Sci.* **8**, 239–241 (2004).
- Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J. & Frith, C. D. Brain responses to the acquired moral status of faces. *Neuron* **41**, 653–662 (2004).
- Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
- Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425**, 785–791 (2003).
- Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* **100**, 3531–3535 (2003).
- Cardinal, R. N., Parkinson, J. A., Hall, J. & Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* **26**, 321–352 (2002).
- O'Doherty, J. *et al.* Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**, 452–454 (2004).
- Schultz, W. Multiple reward signals in the brain. *Nature Rev. Neurosci.* **1**, 199–207 (2000).
- de Quervain, D. J. *et al.* The neural basis of altruistic punishment. *Science* **305**, 1254–1258 (2004).
- Davis, M. A. A multidimensional approach to individual differences in empathy. *JSAS Cat. Selected Docs Psychol.* **10**, 85 (1980).
- Dayan, P. & Balleine, B. W. Reward, motivation, and reinforcement learning. *Neuron* **36**, 285–298 (2002).
- Montague, P. R., Hyman, S. E. & Cohen, J. D. Computational roles for dopamine in behavioural control. *Nature* **431**, 760–767 (2004).
- Berg, J., Dickhaut, J. & McCabe, K. Trust, reciprocity and social history. *Games Econ. Behav.* **10**, 122–142 (1995).
- Friston, K. J. *et al.* Spatial registration and normalization of images. *Hum. Brain Mapp.* **2**, 165–189 (1995).
- Josephs, O., Turner, R. & Friston, K. Event-related fMRI. *Hum. Brain Mapp.* **5**, 243–248 (1997).
- Friston, K. J. *et al.* Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **2**, 189–210 (1995).
- Penny, W. & Holmes, A. P. in *Human Brain Function* (eds Frackowiak, R. S. J. *et al.*) 843–850 (Elsevier, San Diego, 2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Aston, S. Kiebel and E. Featherstone for their help. This work was supported by the German Academy of Natural Sciences Leopoldina, Halle, from the Ministry of Education and Science, by the Medical Research Council (UK), and by Wellcome Trust Programme Grants to R.J.D. and C.D.F.

Author Contributions T.S. designed, implemented, analysed and wrote the paper. B.S., K.E.S. and J.P.O. helped with scanning. All authors contributed to designing the study, discussing the data and preparing the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.S. (t.singer@fil.ion.ucl.ac.uk).

LETTERS

The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells

Benjamin Ohlstein¹ & Allan Spradling¹

Vertebrate and invertebrate digestive systems show extensive similarities in their development, cellular makeup and genetic control¹. The *Drosophila* midgut is typical: enterocytes make up the majority of the intestinal epithelial monolayer, but are interspersed with hormone-producing enteroendocrine cells^{2–4}. Human (and mouse) intestinal cells are continuously replenished by stem cells, the misregulation of which may underlie some common digestive diseases and cancer⁵. In contrast, stem cells have not been described in the intestines of flies, and *Drosophila* intestinal cells have been thought to be relatively stable^{6–7}. Here we use lineage labelling to show that adult *Drosophila* posterior midgut cells are continuously replenished by a distinctive population of intestinal stem cells (ISCs). As in vertebrates, ISCs are multipotent, and Notch signalling is required to produce an appropriate fraction of enteroendocrine cells⁸. Notch is also required for the differentiation of ISC daughter cells, a role that has not been addressed in vertebrates. Unlike previously characterized stem cells, which reside in niches containing a specific partner stromal cell^{9,10}, ISCs adjoin only the basement membrane, differentiated enterocytes and their most recent daughters. The identification of *Drosophila* intestinal stem cells with striking similarities to their vertebrate counterparts will facilitate the genetic analysis of normal and abnormal intestinal function.

Encouraged by histological studies describing diploid cells in the *Drosophila* midgut^{2,11} that morphologically resemble the 'regenerative cells' of beetle, cockroach and moth intestine^{12–14}, we searched the adult *Drosophila* gut for the presence of stem cells. Focusing on the posterior midgut (Fig. 1a), we first identified useful antibodies to mark major intestinal cell types. Interspersed among the monolayer of enterocytes are the enteroendocrine cells, which can be identified by their expression of peptides such as Allatostatin³ and Tachykinin⁴ (Fig. 1b). We found that midgut enteroendocrine cells usually occur in pairs expressing different peptides, and that all known enteroendocrine cells specifically express the transcription factor Prospero¹⁵ (Fig. 1b, c). Staining for the β -Catenin homologue Armadillo¹⁶ outlines nests of 2–3 basally located diploid cells that lack Prospero (Fig. 1d, arrowhead) as well as polyploid enterocytes, with their large, medially located nuclei (Fig. 1d, arrow), and Prospero-positive enteroendocrine cells with their small nuclei (Fig. 1d, asterisk). These cell nests are readily recognizable in electron micrographs (Fig. 1e). One of the cells in each group extensively contacts the basement membrane, and its partner is more apical. Both are surrounded on all sides by enterocytes, which close over the top, sometimes creating a surface depression. These small clusters correspond to the 'regenerative cells' described previously^{2,11}.

To demonstrate that the cell nests contain a functional stem cell, we used a sensitive lineage labelling method¹⁷ that has identified other *Drosophila* stem cells^{18,19}. We permanently marked individual cells

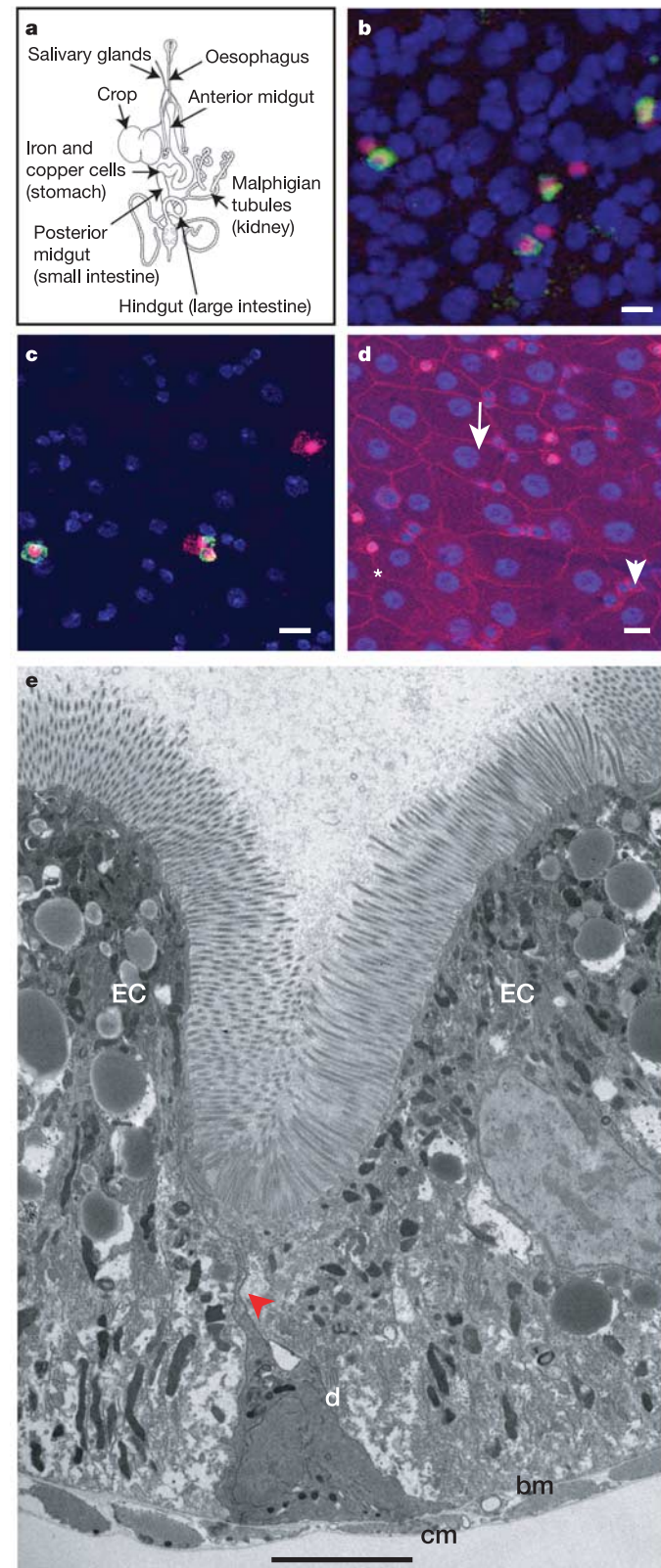
and their descendants by briefly heating 2-, 5- or 10-day-old adult females bearing a heat-shock-inducible (hs)-*FLP* transgene and two inactive *lacZ* constructs (Fig. 2a) to activate a functional gene producing a nuclear-targeted β -galactosidase (β -gal) protein in a small number of random, widely separated dividing cells. At various times after heat-shock, we stained with specific antibodies for β -gal, Armadillo and Prospero in order to follow the growth and differentiation of the marked cells. Marking a non-stem cell produces a 'transient' clone with limited proliferation potential, whereas stem cell clones, unless lost, can divide indefinitely. We found that both types of clones were present (Fig. 2b). Stem cell clones grew in size with increasing time after heat-shock and were surprisingly common. We term the progenitors of these clones 'intestinal stem cells' (ISCs). Transient clones almost always contained only one cell, either a mature enterocyte or, more rarely, an enteroendocrine cell (Fig. 2c). Their unicellular size demonstrates that no mitotic divisions occur in the intestinal lineage beyond the stem cell. Consistent with this conclusion, 49–63% of clones were transient, which is close to the predicted value of 50% and lower than the 75% expected if even a single transit division occurred (Fig. 2d). Moreover, stem cell clone size increased linearly rather than exponentially (Fig. 2e, blue line), with slightly more than one division per day. Thus, the *Drosophila* midgut contains abundant stem cells that follow an extremely simple lineage. Similar results were obtained regardless of the age of the starting flies, as long as nutrients were optimal and free of enteropathogens throughout the experiment. Stem cells were also observed in male midguts and in other gut regions (data not shown), but here we focused on the posterior midgut because of its cellular simplicity, and on females because of the high nutritional demand of egg production²⁰.

These studies also demonstrate that the adult *Drosophila* midgut undergoes rapid cell turnover, like the vertebrate intestine. Starting three days after heat-shock, when transient clones can be recognized unambiguously (ISC clones would be larger by this time), their number is seen to decrease rapidly (Fig. 2e, red line), with turnover virtually complete within one week. Staining with a marker for apoptosis revealed positive cells scattered throughout the tissue (Fig. 2f). The nuclei of these cells also usually showed morphologic evidence of apoptosis. The presence of dispersed apoptotic cells suggests that the *Drosophila* midgut loses cells in a similar manner to many other insects—by delamination of individual apoptotic midgut cells into the gut lumen^{11–13}. An actin purse-string mechanism is proposed to maintain the integrity of the epithelium during this process²¹. Our data indicate that under conditions of active digestion, *Drosophila* intestinal cells turn over approximately weekly and are continuously replaced by the progeny of ISCs.

The structure of ISC stem cell clones shows that they correspond to the basal diploid cells of the cell nests, and provides insight into their regulation and behaviour. By observing ISC clones of increasing size

¹Howard Hughes Medical Institute Research Laboratories, Department of Embryology, Carnegie Institution of Washington, 3520 San Martin Drive, Baltimore, Maryland 21218, USA.

(Fig. 3a, b) and the location of transient clones (Fig. 2b), it is clear that clonal expansion occurs from the basal diploid cell upwards (Fig. 3a, dashed arrow). Thus, the most basal cell corresponds to the ISC, and it contacts its immature diploid daughter, a cell we term the 'enteroblast' (EB). EBs do not divide, but they continue to move apically and frequently become displaced to one side,



presumably towards a region where apoptosis has occurred and a new cell is needed. Following the next ISC division, the former EB frequently begins to expand in size and DNA content, indicating development along the enterocyte pathway. Staining for Armadillo revealed high levels on the membrane separating ISCs and EBs, with lower levels on mature enterocytes (Fig. 3a, b). The organization of the ISC and its surrounding cells is summarized in Fig. 3c.

Vertebrate intestinal stem cells are multipotent, each giving rise to enterocytes, secretory cells and more than ten different types of enteroendocrine cells⁵. *Drosophila* posterior midguts also contain two (and probably more) types of enteroendocrine cells arranged in heterologous pairs (Fig. 1c, d), in addition to enterocytes and possibly other cell types. Staining ISC clones with cell-type markers revealed that individual *Drosophila* ISCs give rise to both enterocytes and at least two types of enteroendocrine cell (Fig. 3d). Marked EBs were observed to differentiate directly into all three cell types, indicating that cell fate is probably selected at this stage. The two members of an enteroendocrine cell pair do not arise from a common precursor, because clones often contained only one pair member, and transient clones consisting of both pair members were almost never observed. These experiments show that ISCs are multipotent stem cells, like their vertebrate counterparts.

The Notch signalling pathway controls the production of secretory cell types downstream of the stem cell in the mammalian intestine⁶. Mice mutant for components of this pathway show greatly increased numbers of enteroendocrine cells. We found that *Drosophila Notch* is strongly expressed in the early cells of the ISC lineage, including the stem cell, EB and young differentiating enterocytes, but not in mature enterocytes or enteroendocrine cells (Fig. 4a). When flies bearing a temperature-sensitive allele of *Notch* (*Notch^{ts}*) (ref. 22) were raised at the non-permissive temperature, the number of midgut enteroendocrine cells increased greatly (Fig. 4b). Moreover, when flies were treated with the γ -secretase inhibitor DAPT, which blocks processing of both vertebrate and *Drosophila* Notch²³, large clusters of Prospero-positive cells accumulated in the midgut (Fig. 4c).

To determine whether Notch signalling is directly required within the ISC lineage, we generated green fluorescent protein (GFP)-marked clones of the null *Notch^{55e11}* allele. Two types of mutant clones were observed in approximately equal numbers, which we believe is due to the segregation of the mutation into either an ISC or its EB daughter following FLP-mediated recombination. Some clones give rise to growing clusters of diploid cells that express the

Figure 1 | The *Drosophila* posterior midgut and its cells. **a**, A drawing (adapted from ref. 2) of the *Drosophila* gut shows the location of the posterior midgut. Some proposed homologies to the vertebrate digestive system are indicated in parentheses. **b**, Tachykinin-positive enteroendocrine cells (green) are interspersed in the epithelial monolayer of enterocytes (large blue nuclei) that lines the posterior midgut. All Tachykinin-positive cells also stain for nuclear Prospero (red). **c**, Simultaneous labelling with Tachykinin (green), Allatostatin (cytoplasmic; red) and Prospero (nuclear; red) shows that Prospero marks all known enteroendocrine cells. All Allatostatin-positive cells were also positive for Prospero. Some Prospero-positive cells not labelled for either Allatostatin or Tachykinin were also present (data not shown). DAPI-stained nuclei are blue. **d**, Low-magnification view of the adult posterior midgut stained to reveal diploid cell nests (arrowhead). Enterocytes are mostly $8n$ and are recognizable by their increased size and DNA content (arrow). Enteroendocrine cells (asterisk) are marked with nuclear Prospero (red). All cells are outlined with membrane-enriched Armadillo (red). **e**, An electron micrograph showing a basal diploid cell (d), including its apical extension (red arrowhead), within the monolayer of polyploid enterocytes (EC). Note that only one diploid cell is visible in this section. A basement membrane (bm) separates diploid nest cells from underlying circular muscle (cm). Scale bar, 10 μ m (**b-d**), 5 μ m (**e**).

neuroendocrine marker Prospero (Fig. 4d). This observation shows that as in mammals, Notch acts early in the intestinal cell lineage to limit the production of enteroendocrine cells. Prospero-positive clones probably result from loss of *Notch* in EBs, followed by differentiation towards the enteroendocrine fate. The remaining clones, which probably arise when *Notch* is lost in ISCs themselves, give rise to growing clusters of diploid cells that do not express Prospero (Fig. 4e). In the absence of *Notch*, ISCs might be unable to differentiate into EBs, causing ISC/EB-like cells to continue proliferating. Both Prospero-positive and Prospero-negative clones grow much more rapidly than a wild-type ISC clone (Fig. 2e, blue), indicating that in the absence of *Notch*, cells downstream of the ISC continue to divide, rather than exiting the cell cycle or becoming polyploid. It would be interesting to determine whether Notch functions similarly in vertebrate ISCs. A model for the roles of Notch in controlling proliferation and differentiation of intestinal cells downstream of the ISC is presented in Fig. 4f.

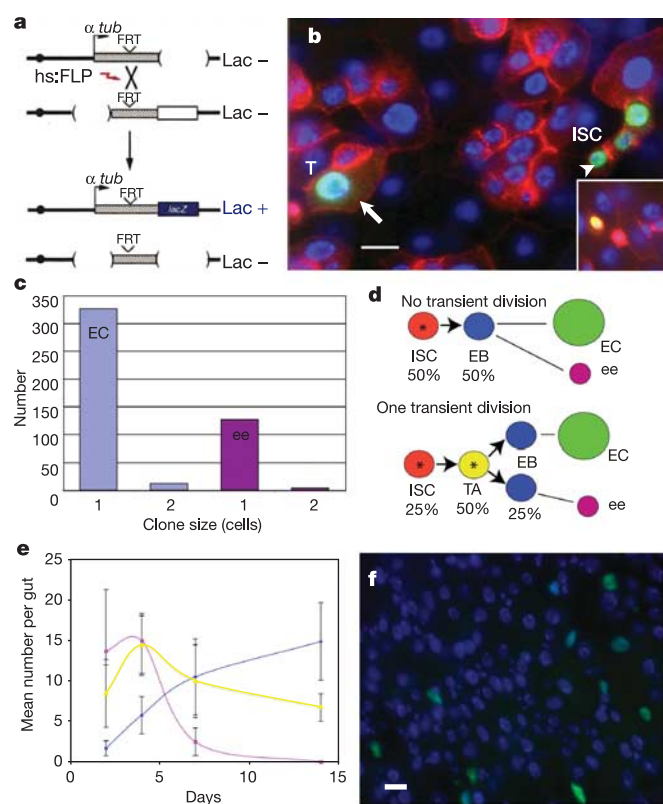


Figure 2 | The midgut contains intestinal stem cells (ISCs). **a**, Scheme¹⁷ for lineage-marking random dividing cells by heat-shock FLP-catalysed site-specific recombination (see Methods). **b**, Survey micrograph showing a stem cell clone (ISC, arrowhead), a transient enterocyte clone (T, arrow) and a transient enteroendocrine cell clone (inset, nuclear Prospero in red). All three clones are also marked with nuclear β -gal (green), Armadillo (red, cytoplasmic) and DAPI (blue). **c**, Graph showing that most transient enterocyte (EC) and enteroendocrine (ee) cell clones contain just one cell. Even the few clones scored as containing two cells may represent independently induced single-cell clones that happen to lie in close proximity. **d**, The predicted fraction of transient clones depends on ISC lineage. Only dividing cells (asterisks) can undergo recombination, and the marker gene can segregate to either daughter. This allows the fraction (shown as a percentage) of ISC and transient clones predicted by each lineage to be calculated. **e**, Kinetics of cell growth and decay. The average size of stem cell clones grows linearly (blue). Transient clones (reflected in the average number per gut) turn over within about one week (red). In contrast, stem cell clones remain stable or decline only slowly (yellow). Error bars indicate s.d. **f**, Apoptotic cells are interspersed throughout the gut, as indicated by ApopTag staining (green). DAPI staining in blue. Scale bar, 10 μ m.

Our findings reveal fundamental similarities between *Drosophila* and vertebrate intestines and their stem cells. Counting the number of cell nests in a gut region of known size allows us to estimate that about 800–1,000 ISCs are dispersed among approximately 10,000 posterior midgut cells, each (as in mammals) maintaining a small local region of the intestine. The general absence of transit divisions downstream of the ISC probably reflects the much smaller size of the *Drosophila* gut compared to the vertebrate gut. This cellular simplicity will facilitate studies of conserved physiological processes, such how enteroendocrine cells, which mediate the contraction of intestinal muscle^{24,25}, are positioned in appropriate numbers adjacent to the gut musculature. The ability to construct genetic mosaics should help to identify and manipulate genes controlling intestinal stem cells and many other aspects of gut function.

Delineation of an insect intestinal stem cell also has important implications for parasitology. Many parasites, including malaria, are transmitted by flies and reside for a time in their guts. Understanding how parasites develop within and escape from the gut of vector insects remains an important issue in designing intervention strategies. Studying the dynamic responses of vector intestinal stem cells to infection will hopefully shed new light on these complex interactions. Moreover, the existence of dividing stem cells suggests potential avenues for intervention.

Finally, the *Drosophila* ISC represents a new model for understanding the fundamental biological mechanisms that control stem cell behaviour. In all of the well-characterized niches, stem cells interact tightly with a non-dividing partner cell that fixes their anatomical location and has an important role in niche function^{9–10}. *Drosophila* ISCs may be controlled differently, because an analogous niche partner does not seem to be present. Instead, ISCs form Armadillo-rich junctions with their own daughter enteroblasts. Such a regulatory regime might allow stem cell mobility along the

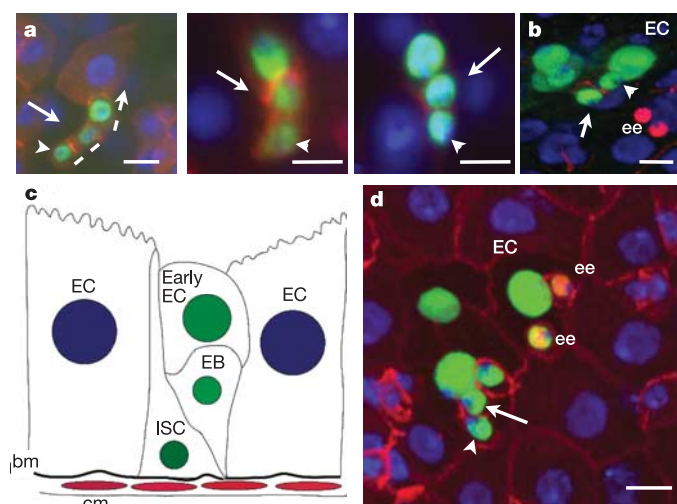


Figure 3 | ISCs reside at the base of the cell nests and are multipotent. **a**, ISC clones have a distinctive morphology and tissue location. Three examples of clones two days after heat-shock show the basal ISC (arrowhead), EB (arrow) and apical direction of growth (dashed arrow). **b**, A six-cell ISC clone (four days after heat-shock) showing the presence of a single ISC and EB, as well as two developing and two mature enterocytes (EC). Two enteroendocrine cells (ee) lie outside the clone. The location of the labelled enterocytes illustrates the movement of ISC progeny cells. **c**, Drawing of a three-cell ISC clone (green nuclei) and its surroundings (compare with **a**). bm, basement membrane; cm, circular muscle. **d**, An eight-cell ISC clone (four days after heat-shock) showing a single ISC, EB, exactly two enteroendocrine cells (ee), and one early and three mature enterocytes (EC). In **a**, **b**, **d**, β -gal (green), Prospero (red, nuclear) and Armadillo (red, cytoplasmic) staining is shown; ISC is indicated by an arrowhead, and EB by an arrow. Scale bar, 10 μ m.

basement membrane. The location of ISCs throughout the midgut is not completely regular, suggesting that they may move in response to changes in the local demand for their cellular products. Clearly, further study of *Drosophila* ISCs should enhance our understanding of stem cells and the niches that control them.

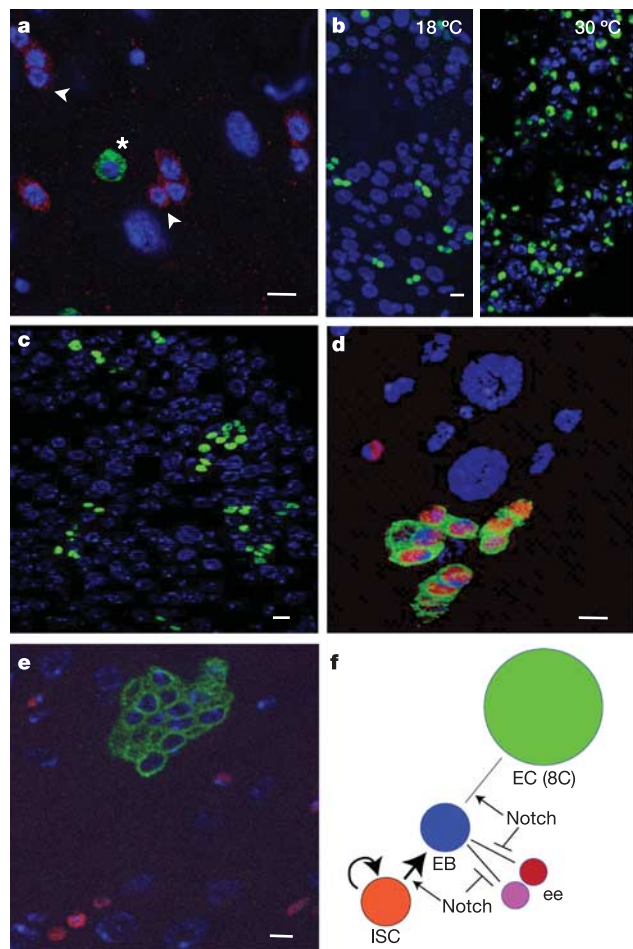


Figure 4 | Notch signalling regulates ISC differentiation. **a**, Notch is preferentially expressed in ISCs, EBs and developing enterocytes. Two ISCs (arrowheads) are visible at the base of developing cell clusters, all of which express Notch (red). Differentiated enteroendocrine cells (asterisk) and mature enterocytes (large nuclei) lack detectable Notch. Tachykinin (green) and DAPI (blue) staining is also shown. **b**, Midgut from *Notch^{ts}* adult at the permissive temperature of 18°C, showing a normal number of enteroendocrine cells (green nuclei). Three weeks after shifting to the restrictive temperature of 30°C, enteroendocrine cells are greatly increased in number and are disorganized. Prospero staining shown in green in both panels. **c**, Proliferation of enteroendocrine-like cells in the presence of the Notch inhibitor DAPT. Prospero staining shown in green. None of the expanded cells stain positively for Tachykinin, suggesting that differentiation of the Prospero-positive enteroendocrine-like cells is arrested (data not shown). **d**, Notch represses endocrine differentiation. A GFP-marked *Notch* null clone (green) gives rise to a mass of enteroendocrine-like cells that express Prospero (red). These may result from the loss of Notch function within an EB. DAPI staining in blue. **e**, Notch may control EB differentiation. A GFP-marked *Notch* null clone (green) giving rise to a mass of diploid cells that do not express Prospero (red). Loss of Notch function within an ISC may block EB differentiation, leading to proliferation of an ISC-like cell. DAPI staining in blue. **f**, Summary of the ISC lineage, illustrating the proposed requirements of Notch in EB differentiation, cell fate choice and cell-cycle exit. EC(8C) indicates mature, octoploid enterocytes. Large arrows indicate cell division. Thin lines indicate differentiation. Positive (small arrows) and negative regulation (T) of steps regulated by Notch are indicated. Scale bar, 10 µm.

METHODS

Drosophila stocks and nomenclature. *Drosophila* stocks were maintained at 21–23°C. A *yw* stock was used as a common genetic background. The temperature-sensitive *Notch* allele has been described²². Further information on genes and symbols can be found in Flybase (<http://flybase.bio.indiana.edu>). The terms ‘intestinal stem cell (ISC)’ and ‘enteroblast (EB)’ are proposed to denote *Drosophila* midgut stem cells and their daughters, respectively (in this paper and ref. 26).

Clonal analysis. Dividing intestinal cells were clonally labelled as described previously¹⁷. Adults of the genotype *yw*hs-FLP/+; *X-15-33*/*X-15-29* were produced by standard crosses. FLP-induced recombination is limited by host factors to dividing cells, and will cause one of the two daughter cells to permanently inherit a gene expressing a nuclear-targeted *Escherichia coli* β-galactosidase. To induce FLP expression, adult flies 2 or 5 days after eclosion (or 10 days in one experiment) were heat-shocked at 38°C for 40 min. No lacZ-positive cells could be found in the intestine in the absence of heat-shock. Each day thereafter, flies were transferred to fresh, yeasted food to stimulate uniform, active digestion during the course of the experiment. Samples of flies were analysed at various times after heat-shock by staining dissected guts with anti-β-galactosidase (β-gal), anti-Prospero and anti-Armadillo antibodies, staining with DAPI and analysing by fluorescence microscopy. In some cases, other antibody combinations were used as described.

Immunostaining and fluorescence microscopy. Intact guts were dissected, fixed in 100 mM glutamic acid, 25 mM KCl, 20 mM MgSO₄, 4 mM sodium phosphate, 1 mM MgCl₂, 4% EM-grade formaldehyde (TedPella) for 30 min and then stained and mounted as described for ovaries²⁷. The following antisera were used: mouse monoclonal anti-Allatostatin³ (1:10), anti-Prospero¹⁵ (1:20), anti-Armadillo¹⁶ (1:20) and anti-Notch-intracellular (1:10) (Developmental Studies Hybridoma Bank); rabbit polyclonal anti-Tachykinin⁴ (1:2,000) (a gift from D. Nässel); rabbit polyclonal anti-β-gal (1:2,000) (Cappel) and rabbit polyclonal anti-GFP (1:5,000) (Torrey Pines Biolabs). Secondary antibodies used were goat anti-mouse and goat anti-rabbit IgG conjugated to either Alexa-488 or Alexa-568 (1:1,000) (Molecular Probes). DAPI (Sigma) was used at 1 µg ml⁻¹. Preparations were examined using a Zeiss Axioimager Z1 equipped with an Apotome system, or a Leica TCS-NT confocal microscope. Ploidy was estimated by quantifying the intensity of DAPI-stained nuclei²⁸. Ovaries were mounted on the same slides as guts, to provide follicle cell nuclei of known ploidy levels.

Analysis of Notch signalling. One-day-old *Notch^{Ax-ts1}* flies were placed at 18°C or 30°C, with daily food changes, and examined as described in the text. One-day-old *yw* flies were fed every two days with N-[N-(3,5-difluorophenacetyl)-L-alanyl]-L-(S)-phenylglycine *t*-butyl ester (DAPT, Sigma) mixed with yeast paste at a final concentration of 2 mM, and examined as described in the text. Clones of the null *Notch^{55c11}* allele were generated using the MARCM system that positively labels homozygous mutant clones with GFP following FLP-catalysed recombination²⁹.

Apoptosis. Apoptotic cells in midgut tissue were identified by staining with Apoptag as described previously²⁰.

Electron microscopy. Intact guts were dissected in Grace's media, then fixed for 1 h at 4°C in 1% glutaraldehyde, 1% OsO₄, 2 mM CaCl₂, 0.1 M cacodylate buffer, pH 7.5. The tissue was then processed and embedded as for ovaries²⁸. Images were captured with a Phillips Tecnai 12 microscope and recorded with a Gatan multiscan CCD camera using Digital Micrograph software.

Received 19 August; accepted 22 September 2005.

Published online 7 December 2005.

1. Stainier, D. Y. No organ left behind: tales of gut development and evolution. *Science* 307, 1902–1904 (2005).
2. Miller, A. in *Biology of Drosophila* (ed. Demerec, M.) 420–534 (Hafner, New York, 1950).
3. Yoon, J. G. & Stay, B. Immunocytochemical localization of *Diploptera punctata* allatostatin-like peptide in *Drosophila melanogaster*. *J. Comp. Neurol.* 363, 475–488 (1995).
4. Siviter, R. J. et al. Expression and functional characterization of a *Drosophila* neuropeptide precursor with homology to mammalian prepro-tachykinin A. *J. Biol. Chem.* 275, 23273–23280 (2000).
5. Radtke, F. & Clevers, H. Self-renewal and cancer of the gut: two sides of a coin. *Science* 307, 1904–1909 (2005).
6. Gartner, L. P. Submicroscopic morphology of the adult *Drosophila* midgut. *J. Baltimore Coll. Dent. Surg.* 25, 64–76 (1970).
7. Gartner, L. P. Ultrastructural examination of ageing and radiation-induced lifespan shortening in adult *Drosophila* midgut. *Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med.* 23, 23–39 (1973).
8. Schonhoff, S. E., Giel-Moloney, M. & Leiter, A. B. Minireview: Development and differentiation of gut endocrine cells. *Endocrinology* 145, 2639–2644 (2004).
9. Fuchs, E., Tumber, T. & Guasch, G. Socializing with the neighbors: stem cells and their niche. *Cell* 116, 769–778 (2004).

10. Ohlstein, B., Kai, T., Decotto, E. & Spradling, A. C. Stem cell niches: theme and variations. *Curr. Opin. Cell Biol.* **16**, 693–699 (2004).
 11. Baumann, O. Posterior midgut epithelial cells differ in their organization of the membrane skeleton from other *Drosophila* epidermis. *Exp. Cell Res.* **270**, 176–187 (2001).
 12. Snodgrass, R. E. *Principles of Insect Morphology* (McGraw-Hill, New York, 1935).
 13. Day, M. F. & Powning, R. F. A study of the processes of digestion in certain insects. *Aust. J. Sci. Res. B* **2**, 175–215 (1949).
 14. Baldwin, K. M. & Kakim, R. S. Growth and differentiation of the larval midgut epithelium during molting. *Tissue Cell* **23**, 411–422 (1991).
 15. Campbell, G. et al. RK2, a glial-specific homeodomain protein required for embryonic nerve cord condensation and viability in *Drosophila*. *Development* **120**, 2957–2966 (1994).
 16. Riggelman, B., Schedl, P. & Wieschaus, E. Spatial expression of the *Drosophila* segment polarity gene *armadillo* is posttranscriptionally regulated by *wingless*. *Cell* **63**, 549–560 (1990).
 17. Harrison, D. A. & Perrimon, N. Simple and efficient generation of marked clones in *Drosophila*. *Curr. Biol.* **3**, 424–433 (1993).
 18. Margolis, J. & Spradling, A. Identification and behaviour of epithelial stem cells in the *Drosophila* ovary. *Development* **121**, 3797–3807 (1995).
 19. Decotto, E. & Spradling, A. C. The *Drosophila* ovarian and testis stem cell niches: similar somatic stem cells and signals. *Dev. Cell* **9**, 501–510 (2005).
 20. Drummond-Barbosa, D. & Spradling, A. C. Stem cells and their progeny respond to nutritional changes during *Drosophila* oogenesis. *Dev. Biol.* **231**, 265–278 (2001).
 21. Han, Y. S., Thompson, J., Kafatos, F. C. & Brillas-Mury, C. Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes. *EMBO J.* **19**, 6030–6040 (2000).
 22. Shellenbarger, D. & Mohler, D. Temperature-sensitive mutations of the Notch locus in *Drosophila melanogaster*. *Genetics* **81**, 143–162 (1975).
 23. Micchelli, C. A. et al. γ -secretase/presenilin inhibitors for Alzheimer's disease phenocopy *Notch* mutations in *Drosophila*. *FASEB J.* **17**, 79–81 (2003).
 24. Nässel, D. R. Tachykinin-related peptides in invertebrates: a review. *Peptides* **20**, 141–158 (1999).
 25. Bendena, W. G., Donly, B. C. & Tobe, S. S. Allatostatins: a growing family of neuropeptides with structural and functional diversity. *Ann. NY Acad. Sci.* **897**, 311–329 (1999).
 26. Micchelli, C. A. & Perrimon, N. Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature* advance online publication, 7 December 2005 (doi:10.1038/nature04371).
 27. Cox, R. T. & Spradling, A. C. A Balbiani body and the fusome mediate mitochondrial inheritance during *Drosophila* oogenesis. *Development* **130**, 1579–1590 (2003).
 28. Lilly, M. & Spradling, A. C. The *Drosophila* endocycle is controlled by cyclin E and lacks a checkpoint ensuring S phase completion. *Genes Dev.* **10**, 2514–2526 (1996).
 29. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
- Acknowledgements** M. Sepanski provided skilled assistance with the electron microscopy. We thank M. Buszczak, E. Decotto and T. Nystul for comments on the manuscript. We are grateful to D. Nässel for providing antibodies and K. Irvine for sending *Notch^{55ell}* on a chromosome with FRT19A.
- Author Information** Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.S. (spadling@ciwemb.edu).

Evidence that stem cells reside in the adult *Drosophila* midgut epithelium

Craig A. Micchelli¹ & Norbert Perrimon^{1,2}

Adult stem cells maintain organ systems throughout the course of life and facilitate repair after injury or disease¹. A fundamental property of stem and progenitor cell division is the capacity to retain a proliferative state or generate differentiated daughter cells²; however, little is currently known about signals that regulate the balance between these processes. Here, we characterize a proliferating cellular compartment in the adult *Drosophila* midgut. Using genetic mosaic analysis we demonstrate that differentiated cells in the epithelium arise from a common lineage. Furthermore, we show that reduction of Notch signalling leads to an increase in the number of midgut progenitor cells, whereas activation of the Notch pathway leads to a decrease in proliferation. Thus, the midgut progenitor's default state is proliferation, which is inhibited through the Notch signalling pathway. The ability to identify, manipulate and genetically trace cell lineages in the midgut should lead to the discovery of additional genes that regulate stem and progenitor cell biology in the gastrointestinal tract.

The adult *Drosophila* midgut can be identified on the basis of two anatomical landmarks along the anterior–posterior axis of the gastrointestinal tract: the cardia and pylorus (Supplementary Fig. S1). The inner surface of the midgut is lined with a layer of cells that project into the gut lumen. These cells exhibit apical–basal polarity, as staining for F-actin reveals the presence of a distinct striated border on their luminal surface (Fig. 1a, b, d and Supplementary Fig. S1). This observation is consistent with the suggestion that the midgut is lined by a cellular epithelium^{3,4}.

Wild-type midguts were stained with 4,6-diamidino-2-phenylindole (DAPI) to reveal the distribution of cell nuclei within the tissue. Nuclei of the midgut display a distinct distribution and fall into two main categories (Fig. 1a–d). The most prominent cells lining the midgut contain large oval nuclei that stain strongly with DAPI (Fig. 1a, b, d). These cells exhibit a region of the nucleus that does not stain with DAPI, giving the nucleus a hollow appearance. This unstained region may correspond to the large nucleolus characteristic of differentiated cells. A second population of cells containing small nuclei can be detected at a basal position within the tissue. The small nuclei are distant from the gut lumen and often lie in close apposition to the two layers of overlying visceral muscle that surround the gut (Fig. 1a–c). On the basis of nuclear size, position and morphology two general populations of midgut cells can, therefore, be distinguished.

Previous studies in *Drosophila* have led to conflicting views over the existence of cell proliferation in the adult gastrointestinal tract^{3–6}. Early reports suggested that somatic stem cells were present in the adult because of morphological similarity to certain larval cells and by analogy to different insect species^{3–5}. In contrast, ³H-thymidine labelling experiments detected DNA synthesis in the adult *Drosophila* midgut, but no mitotic figures were observed in a large sample analysed⁶. On the basis of these observations, the authors concluded

that no somatic cell division occurs during the lifetime of *Drosophila*. To distinguish between these possibilities, we used a series of three independent assays to test whether cell proliferation can be detected in the adult midgut. In the first assay we used genetically marked wild-type cell lineages to identify dividing cells. The production of marked clones after mitotic recombination depends upon subsequent cell division and is, therefore, a direct means to assay proliferation. In these experiments, wild-type lineages were positively marked in adult flies using the MARCM system⁷. Mitotic recombination was induced by heat shock and green fluorescent protein (GFP)-marked clones could be detected in the midgut (Fig. 2b). Similar results were obtained when adults were heat shocked up to 10 days after eclosion (data not shown). This suggests that the ability to generate clones is not transient, and probably persists throughout the entire life of the animal.

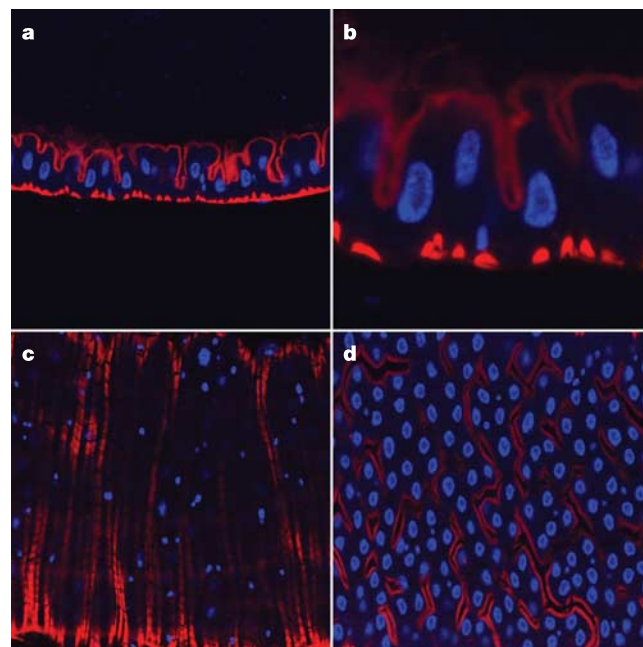


Figure 1 | Cellular organization of the adult *Drosophila* midgut epithelium. **a–d**, Cells with different-sized nuclei occupy distinct positions in the epithelium as seen from two different optical planes (phalloidin, red; DAPI, blue). **a**, A cross-section of the epithelium. The gut lumen is at the top of the image. Magnification is $\times 80$ in each panel, unless otherwise indicated. **b**, High-magnification ($\times 274$). Note the small, basally located nucleus adjacent to the surrounding muscle. **c**, A superficial section of the epithelium. Most small nuclei can be detected at this position. **d**, A deeper luminal section of the epithelium. Most large nuclei can be detected at this position in the epithelium.

¹Department of Genetics, ²Howard Hughes Medical Institute, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

Under the experimental conditions used, the MARCM system produced some background GFP signal that could be detected in control animals (Fig. 2a). To quantify the background signal, we compared the number of GFP-labelled cells in control and experimental animals. A greater than sixfold increase in the number of GFP-labelled cells was detected after heat shock ($n = 10,481$ cells). We also used a second independent clone marking method that did not rely on either Gal4 or Gal80 (ref. 8). In these experiments clones were marked by the loss of a ubiquitously expressed GFP and similar results were observed (data not shown). We conclude that a population of actively dividing somatic cells is present in the adult *Drosophila* midgut.

To extend these findings, we conducted 5-bromodeoxyuridine (BrdU) incorporation studies. We observed both large and small BrdU-labelled midgut cells (Fig. 2c, d). Large nuclei adjacent to each other can be differentially labelled, suggesting asynchrony in the timing or extent of DNA synthesis over the course of the labelling period. This is consistent with the notion that the large nuclei are endoreplicating. However, both endoreplication and the canonical cell cycle require new DNA synthesis. To distinguish endoreplicating

from dividing cells in the midgut we next stained the tissue with an antibody raised against phospho-histone H3. Careful examination revealed that very low levels of phospho-histone H3 staining could be detected in all cells. However, double staining with DAPI revealed that elevated levels of phospho-histone H3 indicative of mitosis could only be detected among the population of cells with small nuclei (Fig. 2e, f). Thus, cells in the midgut seem to have two distinct cell cycles; whereas both large and small nuclei undergo DNA synthesis, only the cells with small nuclei undergo cell division.

In order to characterize further the small cell population, an expression screen was conducted to identify cell-specific molecular markers. We identified three markers expressed in small cells: *escargot* (*esg*), a transcription factor that belongs to the conserved Snail/Slug family; *prospero* (*pros*), a conserved homodomain transcription factor; and *Su(H)GBE-lacZ*, a transcriptional reporter of the Notch signalling pathway⁹. Simultaneous detection of *esg* expression (*esg-Gal4*, *UAS-GFP*), anti-Pros, *Su(H)GBE-lacZ* expression and DAPI demonstrated that small cells could be subdivided into the following classes on the basis of differential gene expression: *esg*-positive (*esg*⁺), *pros*-positive (*pros*⁺), *esg*-negative *pros*-negative (*esg*⁻ *pros*⁻), *esg*-positive

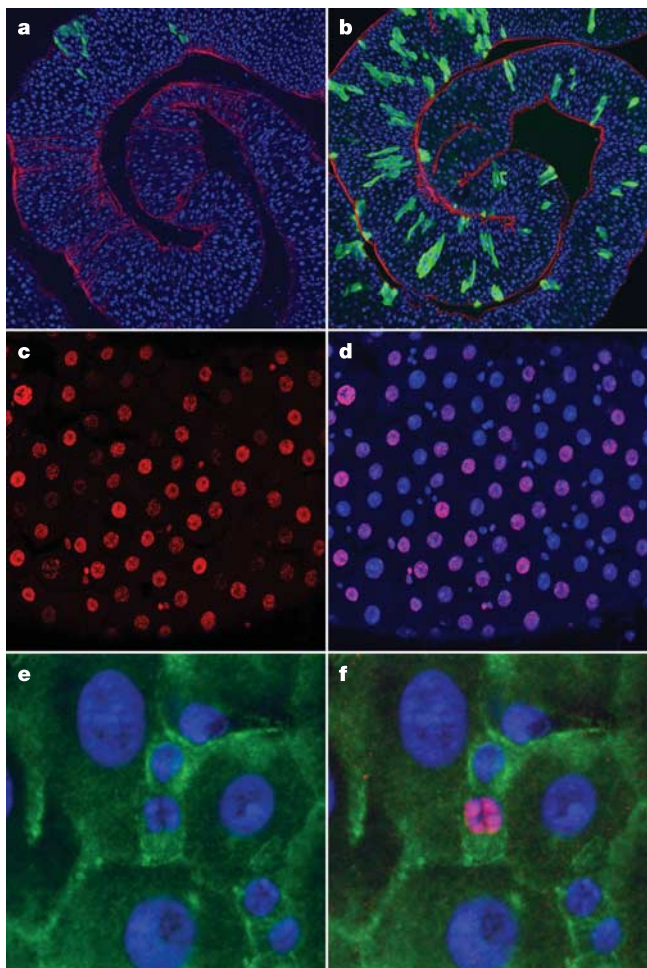


Figure 2 | Cell proliferation in the adult *Drosophila* midgut. **a, b**, Domains of cell proliferation are revealed by the presence of positively marked clones (phalloidin, red; anti-GFP, green; DAPI, blue). Original magnification is $\times 20$. **a**, Controls. No heat shock. **b**, Experimental. Heat shock induction results in the generation of marked clones. **c, d**, DNA synthesis is detected in both large and small cells after a BrdU pulse. **c**, Anti-BrdU. **d**, Overlay (anti-BrdU, red; DAPI, blue). **e, f**, Mitosis is detected in small cells by phospho-histone H3. Original magnification is $\times 320$. **e**, Overlay (*sqh-GFP*, green; DAPI, blue). **f**, Small nucleus with elevated phospho-histone H3 (phospho-histone H3, red; *sqh-GFP*, green; DAPI, blue).

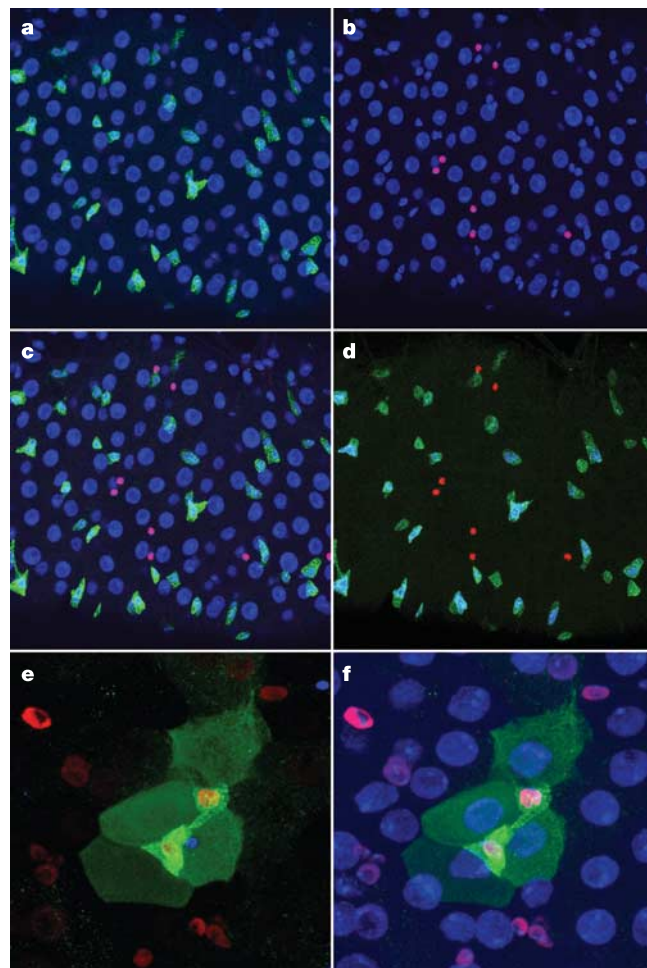


Figure 3 | Cells of the midgut arise from a common lineage. **a–d**, Midgut cells can be distinguished on the basis of differential gene expression. **a**, *esg*⁺ cells (anti-GFP, green; DAPI, blue). **b**, *pros*⁺ cells (anti-Pros, red; DAPI, blue). **c**, *esg*⁺ and *pros*⁺ cells mark distinct populations (anti-Pros, red; anti-GFP, green; DAPI, blue). **d**, *Su(H)GBE-lacZ* expression subdivides the *esg*⁺ cells into two subpopulations (anti-Pros, red; anti-GFP, green; *Su(H)GBE-lacZ*, blue). **e, f**, The cells of the midgut are related by lineage. A positively marked MARCM clone is revealed by anti-GFP staining (green). Original magnification is $\times 160$. **e**, Overlay (*esg-lacZ*, red; anti-GFP, green; anti-Pros, blue). **f**, Overlay (*esg-lacZ*, red; anti-GFP, green; DAPI, blue).

Su(H)GBE-lacZ-positive ($esg^+ Su(H)GBE-lacZ^+$) and *esg*-positive *Su(H)GBE-lacZ*-negative ($esg^+ Su(H)GBE-lacZ^-$) (Fig. 3a–d). esg^+ and *pros*⁺ expression define distinct cell populations (Fig. 3c), whereas *Su(H)GBE-lacZ* expression subdivides the esg^+ class into $esg^+ Su(H)GBE-lacZ^+$ and $esg^+ Su(H)GBE-lacZ^-$ subpopulations (Fig. 3d). Quantification reveals that each cell type is present in the midgut in different proportions (Table 1).

The ability to distinguish different cell types using molecular markers enabled us to determine the cell lineage relationships in this tissue. If the large and small nuclei are lineally distinct then marked clones should be restricted to one or the other cell type. However, if a common stem cell progenitor exists in the adult midgut, then marked lineages should contain both large and small nuclei within a clone. To distinguish between these possibilities we generated positively marked MARCM clones and labelled nuclei using DAPI. Lineage analysis shows that marked clones generated in the adult contain both large and small nuclei (Figs 2b and 3f). In addition, both *esg* expression and anti-Pros-labelled cells could be detected within the clones (Fig. 3e). These lineage-tracing experiments suggest that a stem cell progenitor exists and is sufficient to generate the distinct cell types of the adult midgut. We refer to this cell as the adult intestinal stem cell (ISC).

esg expression in diploid cells has been shown to be necessary for the maintenance of diploidy¹⁰. In addition, the distribution of *esg* messenger RNA has been used as a marker for male germline stem cells¹¹. Together, these observations raise the hypothesis that *esg* expression may also mark a population of progenitors in the midgut. We therefore asked whether *esg* expression correlates with markers of cell proliferation. Simultaneous staining with anti-BrdU and DAPI reveals that *esg*-expressing cells are among the population of cells that are also positively labelled by BrdU (Supplementary Fig. S2a, b). To ask whether *esg*-expressing cells also undergo cell division, the midgut was double stained to detect both *esg* expression and phospho-histone H3. High levels of phospho-histone H3 can be detected specifically in *esg*-expressing cells (Supplementary Fig. S2c). Our results demonstrate that *esg* expression marks a population of proliferating progenitor cells in the midgut.

However, the esg^+ cell population can be divided on the basis of *Su(H)GBE-lacZ* expression (Table 1). To distinguish functionally the two esg^+ populations, we examined the consequences of altering Notch signalling in the adult midgut. We first tested the effect of globally reducing Notch signalling using the conditional *Notch* temperature-sensitive (N^{ts}) mutant. N^{ts} flies were first crossed to an allelic series that included N^{55e11} , $N^{264.47}$, N^{ts1} and $N^{nd.1}$. We found that the strongest loss of function combinations (N^{ts}/N^{55e11} and $N^{ts}/N^{264.47}$) failed to generate viable adult flies even at the permissive temperature, often dying as pharate adults (data not shown). N^{ts}/N^{ts} flies produced viable adults at the permissive temperature with midguts similar to wild type (11 out of 12 midguts; Fig. 4a). N^{ts}/N^{ts} flies shifted to the non-permissive temperature led to a mild increase in the number of small cells (8 out of 12 midguts; Fig. 4b). The weakest allelic combination, $N^{ts}/N^{nd.1}$, also produced viable adults at the permissive temperature but showed no detectable phenotype when shifted to the non-permissive temperature (data not shown).

We next sought to test the requirement of *N* only in esg^+

progenitor cells. To obtain both spatial and temporal control over transgene expression in *esg*-expressing cells, we combined the temperature-sensitive Gal80 inhibitor, *Gal80^{ts}* (ref. 12), with the *esg-Gal4* transcriptional activator. To verify that the *Gal80^{ts}* transgene functions in the midgut, we first characterized the temporal and spatial induction of a *UAS-GFP* transgene. Adult *esg-Gal4, UAS-GFP, tub-Gal80^{ts}* flies grown at the permissive temperature showed no detectable GFP expression in their midguts (Supplementary Fig. S3a, b). In contrast, when these flies were shifted to the non-permissive temperature they showed high levels of GFP expression that were detectable after 1 day and maximal by 2 days (Supplementary Fig. S3c).

The requirement of Notch was then tested in esg^+ cells using a *UAS-N^{RNAi}* transgene¹³, to reduce Notch signalling. In control experiments, *UAS-N^{RNAi}; esg-Gal4, UAS-GFP, tub-Gal80^{ts}* flies grown at the permissive temperature appeared to have wild-type midguts and showed no detectable GFP expression, suggesting that

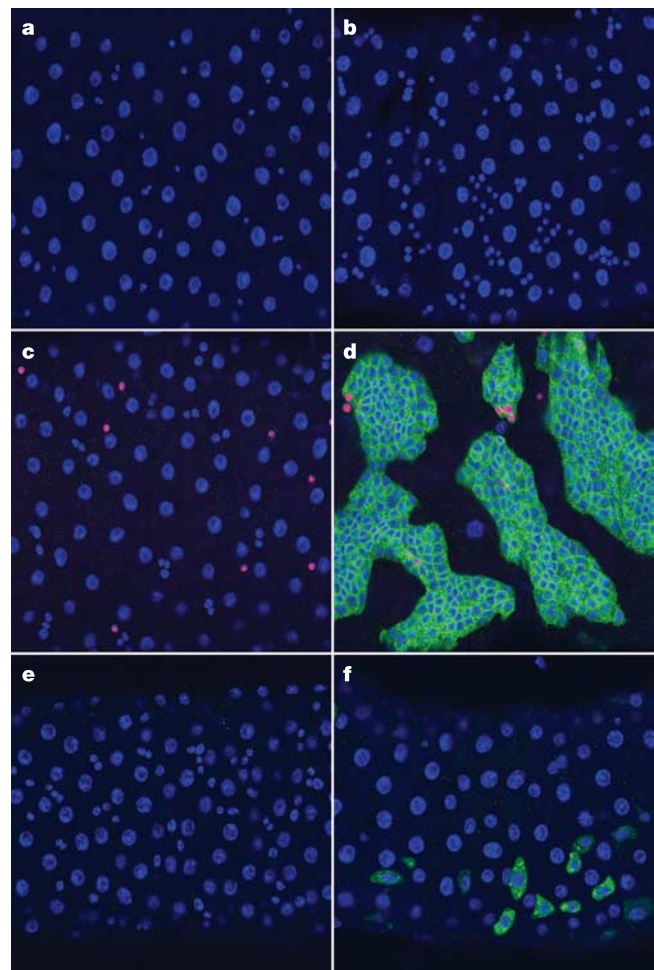


Figure 4 | Notch pathway regulates the number of midgut progenitors. **a, b**, Notch is required to restrict the number of small cells. **a**, Controls. N^{ts} flies grown at the permissive temperature (DAPI, blue). **b**, Experimental. N^{ts} flies shifted to the non-permissive temperature (DAPI, blue). Note the increase in small nuclei. **c, d**, Notch is required in esg^+ progenitors to restrict their number (anti-Pros, red; anti-GFP, green; DAPI, blue). **c**, Controls. *UAS-N^{RNAi}; esg-Gal4, UAS-GFP, tub-Gal80^{ts}* flies grown at the permissive temperature. **d**, Experimental. *UAS-N^{RNAi}; esg-Gal4, UAS-GFP, tub-Gal80^{ts}* flies shifted to the non-permissive temperature. **e, f**, Notch is sufficient to promote the early differentiation of epithelial cells (anti-GFP, green; DAPI, blue). **e**, Controls. *UAS-N^{intra}; esg-Gal4, UAS-GFP, tub-Gal80^{ts}* flies grown at the permissive temperature. **f**, Experimental. *UAS-N^{intra}; esg-Gal4, UAS-GFP, tub-Gal80^{ts}* flies shifted to the non-permissive temperature.

Table 1 | Distribution of small cell types in the adult *Drosophila* midgut

Cell type	Percentage of total in wild type*
esg^+	36
$pros^+$	5
$esg^- pros^-$	1
$esg^+ Su(H)GBE-lacZ^+$	18
$esg^+ Su(H)GBE-lacZ^-$	18

* $n = 242$.

under these conditions *UAS* transgenes are efficiently suppressed (0 out of 20 midguts; Fig. 4c). In contrast, *UAS-N^{RNAi}*; *esg-Gal4*; *UAS-GFP*; *tub-Gal80^{ts}* flies shifted to the non-permissive temperature showed an increase in the number of small cells (19 out of 20 midguts; Fig. 4d). Notably, the presence of *esg-Gal4*, *UAS-GFP* in this experiment enabled us to determine that the increased number of small cells were also *esg⁺*. When these guts were co-stained with anti-Pros antibody we observed ectopic small cells that also expressed *pros*, and these cells were often associated with lower levels of *esg* expression (Fig. 4d). Taken together our experiments suggest that Notch signalling in *esg⁺* cells is necessary to restrict proliferation.

We also tested the effect of Notch activation in *esg⁺* cells using *N^{intra}*, a constitutively active form of Notch. In control experiments, *esg-Gal4*; *UAS-GFP*; *tub-Gal80^{ts}*; *UAS-N^{intra}* flies grown at the permissive appeared to have wild-type midguts and showed no detectable GFP expression (0 out of 23 midguts; Fig. 4e). In contrast, *esg-Gal4*; *UAS-GFP*; *tub-Gal80^{ts}*; *UAS-N^{intra}* flies shifted to the non-permissive temperature showed a decrease in phospho-histone H3 staining compared to controls that were not shifted (mitotic index = phospho-histone H3 positive cells/adult midgut: mitotic index experimental = 0.1, $n = 17$; mitotic index controls = 16, $n = 15$). In addition, although some *esg⁺* cells appear to be wild type, we observed a region-specific decrease in the levels of *esg* expression and a concomitant increase in nuclear size similar to that of midgut epithelial cells (Fig. 4f). These observations demonstrate that Notch activation is sufficient to limit proliferation of *esg⁺* cells and suggests that Notch may also be sufficient to promote early steps of epithelial cell differentiation.

Taken together, our characterization of the adult *Drosophila* midgut suggests that a population of adult stem cells resides within this tissue. We note that this general conclusion of our study is consistent with an accompanying paper¹⁴. Our analysis of the Notch signalling pathway in *esg⁺* cells suggests that *esg⁺ Su(H)GBE-lacZ⁻* cells mark a population of dividing progenitors and that Notch is necessary and sufficient to regulate proliferation. We propose a model in which *esg⁺ Su(H)GBE-lacZ⁻* progenitors generate at least two different types of daughter cells depending on the level of Notch activation (Supplementary Fig. S4a–c). Under conditions of reduced Notch function we observed an expansion of both *esg⁺* progenitor cells and *pros⁺* cells. These observations suggest that *esg⁺* cells give rise to *pros⁺* cells in a Notch-independent manner. Under conditions of Notch activation we observed a decrease in the proliferation and promotion of epithelial cell fate differentiation, while the number of *pros⁺* cells remained unaffected (data not shown).

Several lines of evidence suggest that *pros⁺* cells correspond to gut enteroendocrine cells. Previous studies show that *prox1*, the vertebrate *pros* homologue, is associated with post-mitotic cells and early steps of differentiation in the central nervous system¹⁵. Furthermore, in *Drosophila*, *pros* is thought to be a pan-neural selector gene that is both necessary and sufficient to terminate cell proliferation¹⁶. Finally, although vertebrate enteroendocrine cells arise from endodermal origins they are known to express neural-specific markers¹⁷. Therefore, *pros⁺* cells probably define a population of enteroendocrine cells in the midgut.

Studies of stem cell compartments in *Drosophila* have led to the characterization of two types of progenitor cells in the germ line^{18,19}. The first is referred to as the germline stem cell and is sufficient to give rise to the respective cells of either the male or female germ line. The second type of progenitor cell described is called the cystoblast in female germ line and gonialblast in the male germ line. Although the cystoblast and gonialblast both have the capacity to generate the differentiated cells of their respective tissues, they are thought to be more restricted in their fate than the germline stem cells. On this basis, we suggest that an analogous progenitor may also exist in the adult *Drosophila* midgut; we refer to this cell as the enteroblast (EB). The population of *esg⁺ Su(H)GBE-lacZ⁻* progenitor cells, which we have described, display characteristics of both the ISC and the EB;

therefore, additional experiments will be necessary to distinguish unambiguously these alternatives.

METHODS

Fly culture and strains. Flies were maintained on standard media. Crosses were cultured at either 18 or 25 °C in humidity controlled incubators on a 12 h light/dark cycle. In the experiments described here only female flies were analysed. The following fly strains were used: *Oregon^R* (wild type); *esg-Gal4* (Hayashi laboratory); *UAS-mCD8GFP*; *esg^{K606}/Cyo* (*esg-lacZ*); *y,w, hsf1p¹²²*; *FRT^{40A}, ubiq GFP/Cyo*; *y,w, FRT^{40A} y +*; *y,w, UAS-GFP, hsf1p, tubulin-Gal4*; *FRT^{82B}, tubulin Gal80/TM6B*; *y,w, hsf1p, FRT^{82B}, hs pi-Myc*; *y,w, hsf1p, esg^{K606}/Cyo*; *FRT^{82B}, hs pi-Myc, sqh-GFP*; *P(tubP-Gal80^{ts})9,w/FM7c*; *Su(H)GBE-lacZ* (Bray laboratory); *N^{55c11}* (null allele of *N*); *N^{264.47}*; *N^{ts1}* (a temperature-sensitive allele of *N*); and *N^{md.1}*; *w, UAS-N^{RNAi}*; *w, UAS-N^{intra}* (constitutively active form, deleted extracellular domain). All stocks were obtained from the Bloomington stock centre unless otherwise indicated.

Mosaic analysis. For MARCM clones, fly crosses were established and cultured at 18 °C to generate adults of the genotype *y,w, UAS-GFP, hsf1p, tubulin-Gal4/+*; *FRT^{82B}, tubulin Gal80/FRT^{82B}, hs pi-Myc* or *y,w, UAS-GFP, hsf1p, tubulin-Gal4/esg^{K606}, FRT^{82B}, tubulin Gal80/FRT^{82B}, hs pi-Myc*. Negatively marked clone crosses were established and cultured at 25 °C to generate adults of the genotype *y,w, hsf1p¹²²/y, w, FRT^{40A}, ubiq GFP/FRT^{40A} y +*. Equal numbers of adult flies were then divided into experimental and control groups. Experimental animals were subjected to between one and three 37 °C heat shocks for clone induction. After heat shock, MARCM clones were grown at 18 °C to minimize background signal from the Gal4/UAS system. Control flies did not receive heat shock and were maintained at 18 °C. Midguts were examined between 7–27 days after clone induction.

Temperature shift experiments. Crosses were established and cultured at 18 °C, the permissive temperature, until adulthood. The progeny were divided into two equal pools; controls were cultured at 18 °C and the experimental group was shifted to 29 °C. Flies were shifted to 29 °C for the following periods of time: *N^{ts}* (15–18 days); *UAS-N^{RNAi}*; *esg-Gal4*; *UAS-GFP, tub-Gal80^{ts}* (17–19 days); *esg-Gal4*; *UAS-GFP, tub-Gal80^{ts}*; *UAS-N^{intra}* (7–17 days). Cultures were transferred onto fresh food augmented with yeast paste every 2–3 days during the experimental period.

BrdU labelling. BrdU staining was performed using standard methods and modified as follows. Adult flies were cultured on standard fly media vials augmented with 200 µl of 6 mg ml⁻¹ BrdU plus 20% sucrose. Flies were subsequently cultured at 25 °C and transferred to new media every 2 days for between 4–8 days to achieve maximal labelling.

Histology. Adult flies were dissected in *Drosophila* Ringer's solution. The entire gastrointestinal tract was removed and fixed in ×0.5 PBS (Sigma) plus 4% EM grade formaldehyde (Polysciences) for 30 min. Samples were washed in ×1 PBS plus 0.1% Triton X-100 (PBST) for 2 h, then incubated with primary antibodies over night. Primary antibodies were removed and samples were washed in PBST for 2 h. Secondary antibodies were incubated for 3 h. Secondary antibodies were removed and samples washed in PBST for 2 h. Mounting media containing DAPI (Vectashield) was added and the samples were allowed to clear for 1 h before mounting. All steps completed at 4 °C, with no mechanical agitation.

Antisera. The following primary antibodies were used: goat anti-GFP (Molecular Probes) 1:2,000; rabbit anti-β-gal (Cappel) 1:2,000; rabbit anti-phospho-histone H3 (Upstate Biotechnology) 1:1,000; mouse anti-tubulin (Sigma) 1:1,000; mouse anti-BrdU (Becton Dickinson) 1:100; anti-Pros (Developmental Studies Hybridoma Bank) 1:40. The following secondary antibodies were used: donkey anti-goat Alexa 488 (Molecular Probes) 1:500; donkey anti-goat Alexa 594 (Molecular Probes) 1:500; donkey anti-goat Alexa 633 (Molecular Probes) 1:500; donkey anti-rabbit rhodamine RRX (Jackson ImmunoLabs) 1:250; donkey anti-mouse Cy5 (Jackson ImmunoLabs) 1:250. We used the following dyes and mounting media: Alexa 546- and Alexa 594-conjugated phalloidin (Molecular Probes) 1:500; Vectashield plus DAPI (Vector).

Microscopy and imaging. Fluorescent samples were examined on a Zeiss Axioskop 2motplus upright microscope. Confocal images were collected using a Leica TCS SP2 AOBs confocal system. Images were processed and assembled in Photoshop CS.

Nomenclature. In consensus with the accompanying paper¹⁴, we have agreed to use the terms intestinal stem cell (ISC) and enteroblast (EB) to refer to progenitor cells of the adult *Drosophila* midgut epithelium.

Received 19 August; accepted 26 September 2005.

Published online 7 December 2005.

1. Radtke, F. & Clevers, H. Self-renewal and cancer of the gut: Two sides of a coin. *Science* 307, 1904–1909 (2005).

2. Molofsky, A. V., Pardoll, R. & Morrison, S. J. Diverse mechanisms regulate stem cell self-renewal. *Curr. Opin. Cell Biol.* **16**, 700–707 (2004).
3. Strasburger, M. Bau, Funktion und Variabilität des Darmtractus von *Drosophila melanogaster* Meigen. *Zeit. f. wiss. Zool.* **140**, 539–649 (1932).
4. Wigglesworth, V. *The Principles of Insect Physiology* (Methuen & Co. Ltd, London, 1965).
5. Miller, A. in *The Biology of Drosophila* (ed. Demerec, M.) 420–442 (Hafner, New York, 1950).
6. Bozuck, A. N. DNA synthesis in the absence of somatic cell division associated with ageing in *Drosophila subobscura*. *Exp. Geront.* **7**, 147–156 (1972).
7. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neural morphogenesis. *Neuron* **22**, 451–461 (1999).
8. Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).
9. Furriols, M. & Bray, S. J. A model Notch response element detects Suppressor of Hairless dependent molecular switch. *Curr. Biol.* **11**, 60–64 (2001).
10. Fuse, N., Hirose, S. & Hayashi, S. Diploidy of *Drosophila* imaginal cells is maintained by a transcriptional repressor encoded by escargot. *Genes Dev.* **8**, 2270–2281 (1994).
11. Kiger, A., White-Cooper, H. & Fuller, M. T. Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature* **407**, 750–754 (2000).
12. McGuire, S. E. *et al.* Spatiotemporal rescue of memory dysfunction in *Drosophila*. *Science* **302**, 1765–1768 (2003).
13. Presente, A., Shaw, S., Nye, J. S. & Andres, A. J. Transgene-mediated RNA interference defines a novel role for Notch in chemosensory startle behaviour. *Genesis* **34**, 165–169 (2002).
14. Ohlstein, B. & Spradling, A. The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. *Nature* advance online publication, 7 December 2005 (doi:10.1038/nature04333).
15. Torii, M. *et al.* Transcription factors Mash-1 and Prox-1 delineate early steps in differentiation of neural stem cells in the developing central nervous system. *Development* **126**, 443–456 (1999).
16. Li, L. & Vaessin, H. Pan-neural prospero terminates cell proliferation during *Drosophila* neurogenesis. *Genes Dev.* **14**, 147–151 (2000).
17. Schonhoff, S. E., Giel-Moloney, M. & Leiter, A. B. Minireview: Development and differentiation of gut endocrine cells. *Endocrinology* **145**, 2639–2644 (2005).
18. Xie, T. & Spradling, A. in *Stem Cell Biology* (eds Marshak, D. R., Gardner, R. L. & Gottlieb, D.) 129–148 (Cold Spring Harbor Press, New York, 2001).
19. Kiger, A. & Fuller, M. in *Stem Cell Biology* (eds Marshak, D. R., Gardner, R. L. & Gottlieb, D.) 149–187 (Cold Spring Harbor Press, New York, 2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank GETDB and Bloomington stock centers for providing fly strains. C.A.M. would like to acknowledge colleagues in the Perrimon laboratory for many constructive discussions and criticisms; B. Mathey-Prevot for comments on the manuscript; K. Brückner for help in translating the seminal text by M. Strasburger from German; R. Binari for technical assistance; and S. Tang and C. Villata for help in screening. C.A.M. was supported by a NRSA fellowship and a grant from the Harvard Stem Cell Institute. N.P. is a HHMI investigator.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.A.M. (cmichel@genetics.med.harvard.edu).

LETTERS

Pregnenolone stabilizes microtubules and promotes zebrafish embryonic cell movement

Hwei-Jan Hsu^{1,2}, Ming-Ren Liang³, Chao-Tsen Chen³ & Bon-chu Chung¹

Embryonic cell movement is essential for morphogenesis and the establishment of body shapes^{1,2}, but little is known about its mechanism. Here we report that pregnenolone, which is produced from cholesterol by the steroidogenic enzyme Cyp11a1 (cholesterol side-chain cleavage enzyme, P450_{scc})³, functions in promoting cell migration during epiboly. Epiboly is a process in which embryonic cells spread from the animal pole to cover the underlying yolk. During epiboly, *cyp11a1* is expressed in an extra-embryonic yolk syncytial layer⁴. Reducing *cyp11a1* expression in zebrafish using antisense morpholino oligonucleotides did not perturb cell fates, but caused epiboly delay. This epiboly defect was partially rescued by the injection of *cyp11a1* RNA or the supplementation of pregnenolone. We show that the epiboly delay is accompanied by a decrease in the level of polymerized microtubules, and that pregnenolone can rescue this microtubule defect. Our results indicate that pregnenolone preserves microtubule abundance and promotes cell movement during epiboly.

Cyp11a1 (also known as P450_{scc}) catalyses the first step of steroid synthesis, the conversion of cholesterol to pregnenolone (P₅). P₅ is then further converted to 17 α -hydroxypregnenolone by Cyp17 or to progesterone (P₄) by 3 β -hydroxysteroid dehydrogenase (Hsd3b), for the production of other steroids (Fig. 1a). Steroids regulate salt and sugar homeostasis and sexual characteristics in the body, but the functions of steroids during early embryogenesis are unclear. To study their functions, we characterized the genomic structures of zebrafish *cyp11a1* and *hsd3b* (Fig. 1b), and examined their expression during early embryogenesis by *in situ* hybridization. After fertilization, zebrafish blastomeres located above the yolk undergo rapid cleavage. At this time, *cyp11a1* and *hsd3b* were expressed in the cytoplasm of these blastomeres (Fig. 1c). The blastomeres then move downward to cover the yolk in a process called epiboly. During epiboly, *cyp11a1* and *hsd3b* were downregulated from blastomeres, and a new expression domain appeared at the junction between blastomeres and yolk. This expression domain, called the yolk syncytial layer (YSL), is an extra-embryonic layer composed of yolk syncytial nuclei that move ahead of blastomeres during epiboly.

To test the function of *cyp11a1* during early embryogenesis, we injected antisense morpholino (mo) oligonucleotides into zebrafish embryos at the one-cell stage to block its expression (Supplementary Fig. S1). At 100% epiboly, when wild-type blastomeres completely covered the yolk, embryos injected with *cyp11a1*-mo1 could not cover the yolk, leaving an unclosed yolk plug (Fig. 2). Control embryos injected with morpholino oligonucleotides against a random sequence, an unrelated sequence from *dlx3*, *hsd3b* or *cyp17*, or a 5-base mismatched *cyp11a1*-mo1 were all normal (Supplementary Table S1). The epiboly delay started before gastrulation, when wild-type blastomeres cover 50% of the yolk (50% epiboly). At this time, blastomeres of the *cyp11a1*-mo1 morphants covered only about

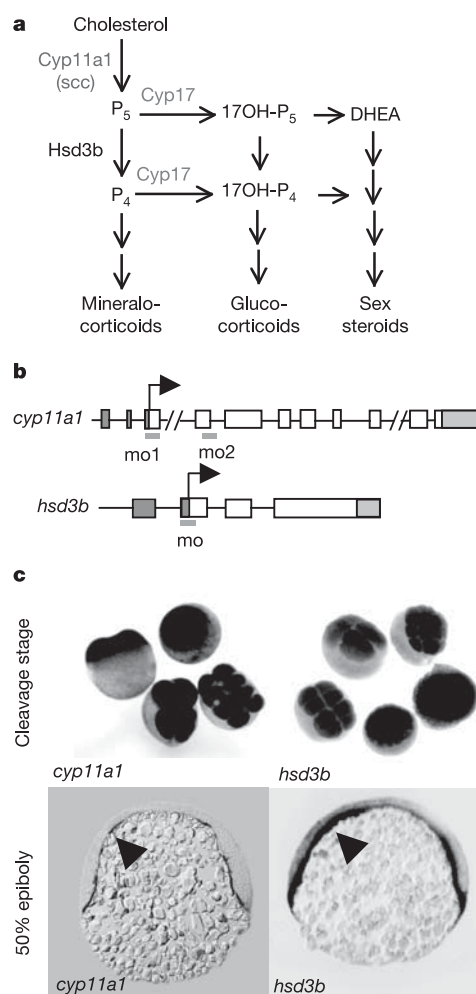


Figure 1 | Two steroidogenic genes, *cyp11a1* and *hsd3b*, are expressed in blastomeres and the yolk syncytial layer during early zebrafish embryogenesis. **a**, Steroidogenic pathways. Cyp11a1 catalyses the conversion of cholesterol to pregnenolone (P₅), which is then converted to 17 α -hydroxypregnenolone (17OH-P₅) and progesterone (P₄) by Cyp17 and Hsd3b, respectively. DHEA, dehydroepiandrosterone. **b**, Genomic structures of zebrafish *cyp11a1* and *hsd3b*. Shaded and open boxes indicate coding and non-coding regions of exons, respectively. Arrows indicate transcriptional start sites; lines below the boxes indicate the locations of morpholino oligonucleotides (mo). **c**, Detection of *cyp11a1* and *hsd3b* expression by whole-mount *in situ* hybridization at the cleavage and 50% epiboly stages. Arrowheads indicate yolk syncytial layer.

¹Institute of Molecular Biology, Academia Sinica, Nankang, Taipei 11529, Taiwan. ²Graduate Institute of Life Sciences, National Defense Medical Center, Taipei 11490, Taiwan.

³Department of Chemistry, National Taiwan University, Taipei 10617, Taiwan.

44% of the yolk (data not shown). Thus, *cyp11a1*-mo1 morphants show an early epibolic defect.

When we injected a fluorescein-conjugated control morpholino oligonucleotide into embryos at the one-cell stage, the fluorescent signal was distributed throughout the entire embryo at 30% epiboly. When we injected the same morpholino into the yolk at the 1,000 (1k)-cell stage, the fluorescent signal was restricted to only the YSL (Supplementary Fig. S2a). Thus, injecting oligonucleotides into the yolk at the 1k-cell stage affects only YSL function. At 100% epiboly, embryos injected with *cyp11a1*-mo1 at the 1k-cell stage failed to close their yolk plug, although the extent of epibolic delay was less severe than that for embryos injected at the one-cell stage (Supplementary Fig. S2b). This result indicates that *cyp11a1* expressed in the YSL participates in the epibolic process.

The YSL is known to be involved in mesoderm induction⁵ and nutrient production^{6,7}. To test whether *cyp11a1* in the YSL is involved in germ layer induction, we examined cell fates in *cyp11a1*-mo1 morphants using mesodermal (*tbx6* and *shh*), mesendodermal (*gsc*) and endodermal (*sox17*) markers. The cell fates of mesoderm and endoderm did not change upon reduction of *cyp11a1* expression by *cyp11a1*-mo1 (Supplementary Fig. S3). Although cell fates of the *cyp11a1*-mo1 morphants appeared normal, the length of their anterior–posterior axis was reduced (Supplementary Fig. S3d) and their somites were wider (data not shown). This indicates a problem with the convergent extension movement that builds up body axis by narrowing and lengthening dorsal tissues. Not all types of cell movements were affected, as involution movement seemed normal (Supplementary Fig. S4).

To confirm the function of *cyp11a1*, we used a different morpholino oligonucleotide, *cyp11a1*-mo2, which targets the junction between exon 4 and intron 4 of *cyp11a1*, and affects zygotic *cyp11a1* RNA splicing and epibolic movement (Fig. 1b and Supplementary Fig. S5). This morpholino also caused epibolic delay. The epibolic delay caused by each of *cyp11a1*-mo1 and *cyp11a1*-mo2 was partially rescued when embryos were co-injected with *cyp11a1* mRNA (Supplementary Figs S2c and S5d). This result shows that *cyp11a1* contributes to epibolic movement.

Cyp11a1 encodes a steroidogenic enzyme (Fig. 1a), and we predicted that reducing *cyp11a1* expression would result in a decrease in steroid products. Indeed, the level of the Cyp11a1 product pregnenolone from *cyp11a1*-mo1 morphants 10 h after fertilization was about half of that in wild-type embryos (Supplementary Fig. S1b), and incubation of *cyp11a1*-mo1 morphants with pregnenolone from the early cleavage stage partially rescued the epibolic defect (Fig. 3a).

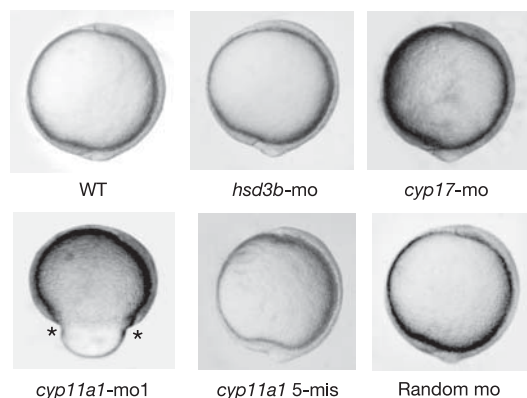


Figure 2 | Reduced *cyp11a1* expression results in epibolic delay.

a, Phenotypic examination of embryos under a dissecting microscope. Wild-type (WT) embryos and embryos injected with control, *cyp11a1*-mo1, *cyp11a1* 5-mis (5-base mismatched *cyp11a1*-mo1), *hsd3b* or *cyp17* morpholino oligonucleotides (mo) at the one-cell stage reach 100% epiboly. Embryos injected with *cyp11a1*-mo1 cannot close the yolk plug.

This indicates that pregnenolone and/or its metabolite is involved in zebrafish epiboly.

To determine whether pregnenolone is metabolized further, we incubated embryos with radioactive pregnenolone and checked the formation of metabolites by thin-layer chromatography. We found that some pregnenolone could be metabolized into progesterone, but that very little 17 α -hydroxypregnenolone was produced, even when Hsd3b activity was knocked down (Supplementary Fig. S6). When tested, progesterone could not rescue the epibolic defect—instead, it worsened the severity of the growth delay (data not shown). As a concentration of only 0.1 μ M pregnenolone was enough to rescue the epibolic defect (Fig. 3a), and this concentration is lower than the K_m for either Cyp17 (5 μ M; ref. 8) or Hsd3b (0.2–1.6 μ M; ref. 9), it is likely that pregnenolone, rather than its metabolite, rescues the epibolic delay.

Zebrafish epiboly is driven by yolk microtubule arrays, which extend from yolk syncytial nuclei along the animal–vegetal axis of the yolk^{10–12}. To investigate whether the epibolic function of *cyp11a1* is related to microtubules, we visualized microtubules by staining them with an antibody against α -tubulin. We observed no significant difference between wild-type and *cyp11a1*-mo1 morphants during epiboly (data not shown). We therefore tested the stability of microtubules by treating embryos at 50% epiboly with the microtubule-disrupting agent nocodazole. While almost completely disrupted by 10 μ g ml⁻¹ nocodazole¹¹ (data not shown), yolk microtubules remained intact with treatment of 1 μ g ml⁻¹ nocodazole, but the yolk microtubules of *cyp11a1*-mo1 morphants were thinner than those of wild-type embryos (Fig. 3b). This result shows that yolk microtubules of *cyp11a1* morphants are more sensitive to nocodazole treatment, indicating that they are less stable^{13,14}. When *cyp11a1*-mo1 morphants were incubated with pregnenolone, yolk microtubule structure was restored (Fig. 3b). Therefore, pregnenolone can stabilize yolk microtubules *in vivo*.

In addition to the imaging study described above, we examined microtubule content using a biochemical method. Protein extracts from wild-type embryos and *cyp11a1*-mo1 morphants at 50% epiboly were centrifuged at 100,000 g to separate polymerized tubulin (pellet) from free tubulin (supernatant). When embryos were treated with the tubulin-polymerizing drug taxol, the amount of polymerized tubulin in the pellet was increased (Fig. 3c). In *cyp11a1*-mo1 morphants, the amount of polymerized tubulin in the pellet was much less than for wild type, indicating a lower abundance of polymerized microtubules. When the extract from *cyp11a1*-mo1 morphants was incubated with pregnenolone, the amount of polymerized tubulin in the pellet increased in a dose-dependent manner (Fig. 3d). This indicates that pregnenolone can increase the amount of polymerized tubulin *in vitro*.

To visualize the interaction between pregnenolone and microtubules, we conjugated the fluorescent dye fluorescein to pregnenolone (Supplementary Fig. S7). The resulting F-P₅ also increased the abundance of polymerized tubulin in *cyp11a1*-mo1 morphant extracts (Supplementary Fig. S8), indicating the fluorescein moiety does not interfere with pregnenolone activity. When incubated with *cyp11a1*-mo1 morphant extract for 30 min, F-P₅ co-localized with polymerized α -tubulin (Fig. 4). When fluorescein alone was incubated with morphant extract, it did not co-localize with α -tubulin and microtubules were not detected. This indicates that pregnenolone binds to microtubules, or to something that associates with microtubules in cell extracts. This binding results in increased microtubule content *in vitro*.

Microtubule dynamics provide an important force for cell movement and neurite outgrowth^{15,16}. However, the regulators of microtubules are largely unknown, except for a few identified proteins such as CLASPs, MAP2c, MACF7 and Op18, which stabilize microtubules in various cell types^{13,17,18}. Here we show that pregnenolone is a regulator of yolk microtubules—it maintains the abundance of yolk microtubules and promotes epibolic movement. Pregnenolone

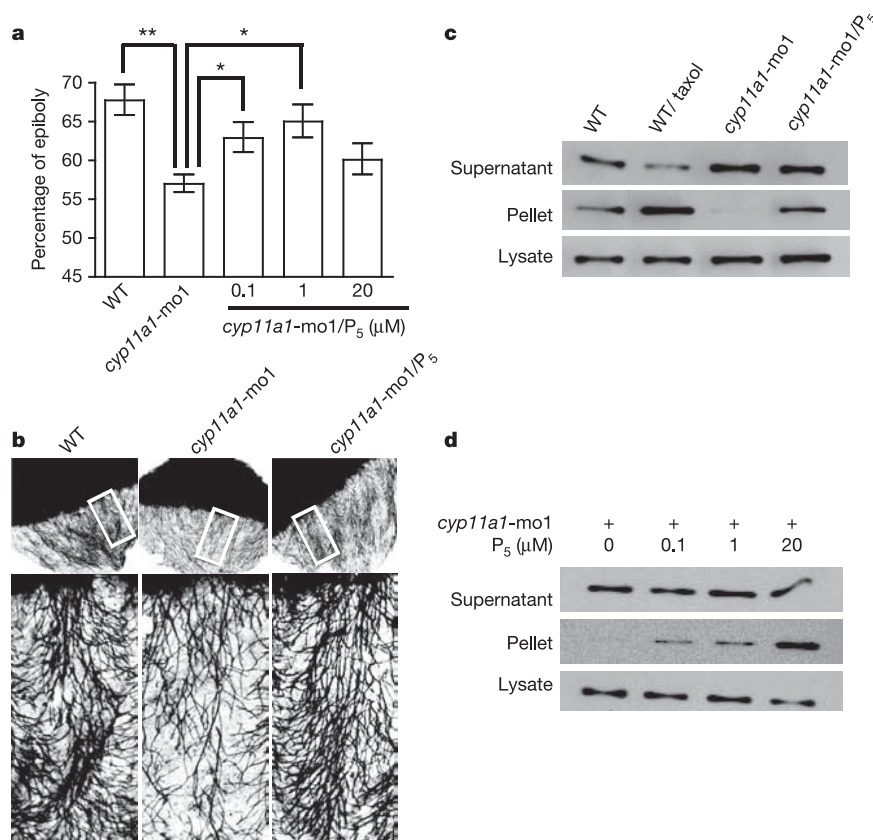


Figure 3 | Pregnenolone rescues epiboly defect and increases microtubule content. **a**, The effect of different concentrations of pregnenolone on the percentage of epiboly. Error bars show s.d. * $P < 0.05$; ** $P < 0.01$. **b**, Microtubule structure of nocodazole-treated wild-type embryos (16 out of 18 embryos showed normal microtubule structure, $n = 16/18$), *cyp11a1-mo1*-injected embryos ($n = 19/23$) or *cyp11a1-mo1* morphants incubated with 20 μM P₅ ($n = 19/19$) at 50% epiboly. Enlarged images show

higher magnifications of the areas in the rectangles. **c**, **d**, Microtubule content assay. Embryo extracts from wild-type (WT) or *cyp11a1-mo1*-injected embryos at 50% epiboly were incubated with or without 20 μM taxol or different concentrations of P₅, centrifuged to separate free tubulin in the supernatant from polymerized microtubules in the pellet, then probed with an α-tubulin antibody. mo, morpholino oligonucleotide.

promotes microtubule assembly through binding to microtubule-associated protein 2 from rat brain¹⁹. This provides insight into the mechanism of cell movement controlled by pregnenolone, not only during zebrafish embryogenesis but probably also for neurite outgrowth and other types of cell movement.

Here we describe a new function for *cyp11a1* expressed in the zebrafish YSL. The YSL is known to have a role in morphogenesis during gastrulation²⁰, but how it participates in morphogenetic cell movement was not clear. We now show that the YSL gene *cyp11a1* is important for the production of pregnenolone, which preserves the abundance of microtubules, thus promoting cell movement.

The production of steroids from extra-embryonic tissues is not unique to zebrafish. In mouse, steroids are secreted from extra-embryonic giant trophoblast cells during early pregnancy²¹, but the function of steroids in mammalian embryogenesis cannot be easily investigated^{22,23}. The *Drosophila* steroid ecdysone is activated in an extra-embryonic tissue and is involved in morphogenetic cell movements for building up the body plan²⁴, but its mechanism of action is not clear. Using zebrafish, we have shown that the steroid pregnenolone, produced from an extra-embryonic tissue, participates in early morphogenesis by acting on microtubules. Steroids usually act through a genomic mechanism, whereby steroids bind to steroid receptors and control the expression of target genes²⁵. Other non-genomic actions have been proposed for steroids²⁶. We now show that pregnenolone promotes cell movements by directly stabilizing microtubules, probably through a non-genomic mechanism.

METHODS

Microinjection and reagent treatment. Morpholino oligonucleotides (Gene Tools) were diluted to working concentrations of 0.5–3 μg μl⁻¹ in Danieau solution (58 mM NaCl, 0.7 mM KCl, 0.4 mM MgSO₄, 0.6 mM Ca(NO₃)₂, 5 mM HEPES, pH 7.6) and injected into the yolk of embryos at the one-cell or 1k-cell stage at a concentration of 9 ng per embryo, unless noted otherwise. For RNA

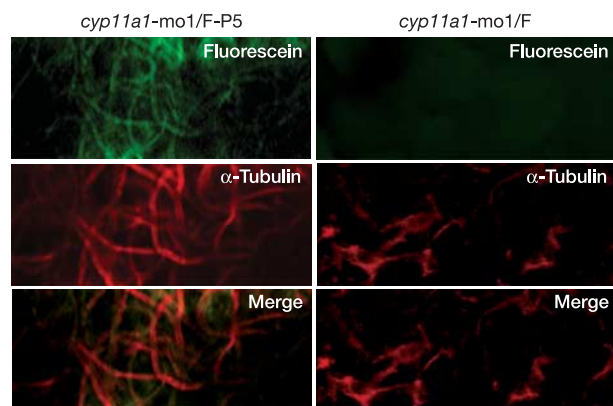


Figure 4 | Fluorescein-conjugated pregnenolone co-localizes with microtubules. Detection of fluorescein and α-tubulin in the microtubule pellet after incubation of *cyp11a1-mo1*-treated embryo extracts with 20 μM taxol and fluorescein-conjugated pregnenolone (F-P₅) for 30 min followed by centrifugation. When incubated with taxol and fluorescein alone (F), no microtubule signal was detected (except for some nonspecific tubulin staining of cell debris).

rescue, 9 ng *cyp11a1*-mo1 and 200 pg of *cyp11a1* RNA or β -gal RNA were co-injected into embryos at the one-cell stage. Embryos were incubated with pregnenolone (0.1, 1 or 20 μ M) or progesterone (1 or 20 μ M) after dechorionation and before the 16-cell stage until the late epiboly stage. For nocodazole treatment, 1 μ g ml⁻¹ nocodazole was added to the embryo medium at 50% epiboly for 20 min at 28.5 °C.

Whole-mount *in situ* hybridization and immunostaining. Whole-mount *in situ* hybridization was performed using digoxigenin-labelled antisense RNA probes and alkaline phosphatase-conjugated secondary antibodies, as described previously²⁷. Photographs were taken using an Olympus BX50 microscope. Microtubule staining with an anti- α -tubulin antibody was performed according to a previously described procedure²⁸. Fluorescent images were captured by Zeiss LSM 510 confocal microscope. Tubulin images were processed by the 3D visualization and deconvolution LSM510 basis software (v3.0, Zeiss).

Measurement of the extent of epiboly. Embryos were hybridized with an *ntl* probe to mark the marginal blastomeres, and photos were taken using a Leica MZFLIII microscope with a CCD camera (Diagnostic Instrument Inc). The length from animal pole to vegetal pole and the height of the blastoderm for each embryo were measured by SPOT program (v.4-0-8, Diagnostic Instrument Inc) (Supplementary Fig. S9). The distance of the blastoderm margin divided by the animal-vegetal length is regarded as the percentage of epiboly.

Synthesis of fluorescein-conjugated pregnenolone (F-P₅). F-P₅ consists of fluorescein and pregnenolone bridged by a six-carbon linker via ethereal and thioetherid functionalities (see Supplementary Information). It was prepared in six steps with 14% overall yield, starting from commercially available pregnenolone.

Microtubule abundance assay. This procedure was according to the instructions provided with the microtubules/tubulin *in vivo* assay kit (Cytoskeleton). Twenty embryos injected with or without 9 ng *cyp11a1*-mo1 were placed in 200 μ l lysis buffer plus 1 mM GTP, 10 mM ATP and 20 μ l proteinase inhibitor cocktail, homogenized, then incubated with taxol (20 μ M), pregnenolone (0.1, 1, 20 μ M), fluorescein or fluorescein-conjugated pregnenolone at 30 °C for 30 min. A 180- μ l aliquot was transferred into a pre-warmed tube and centrifuged at 100,000 g, 30 °C for 30 min. After centrifugation, pellets were dissolved in 180 μ l cold 200 μ M CaCl₂. For western blot analysis, 2.5 μ l each of total lysate and supernatant samples, and 15 μ l of pellet samples were loaded onto the gels. For the imaging study, microtubule pellets were removed from centrifuged tubes and stained with α -tubulin antibody, followed by spotting onto a glass slide for fluorescence detection.

Received 9 September; accepted 25 October 2005.

- Locascio, A. & Nieto, M. A. Cell movements during vertebrate development: integrated tissue behaviour versus individual cell migration. *Curr. Opin. Genet. Dev.* **11**, 464–469 (2001).
- Myers, D. C., Sepich, D. S. & Solnica-Krezel, L. Convergence and extension in vertebrate gastrulae: cell movements according to or in search of identity? *Trends Genet.* **18**, 447–455 (2002).
- Miller, W. L. Molecular biology of steroid hormone synthesis. *Endocr. Rev.* **9**, 295–318 (1988).
- Hsu, H. J., Hsiao, P., Kuo, M. W. & Chung, B. C. Expression of zebrafish *cyp11a1* as a maternal transcript and in yolk syncytial layer. *Gene Expr. Patterns* **2**, 219–222 (2002).
- Chen, S. & Kimelman, D. The role of the yolk syncytial layer in germ layer patterning in zebrafish. *Development* **127**, 4681–4689 (2000).
- Babin, P. J. *et al.* Both apolipoprotein E and A-I genes are present in a nonmammalian vertebrate and are highly expressed during embryonic development. *Proc. Natl. Acad. Sci. USA* **94**, 8622–8627 (1997).
- Sharma, M. K. *et al.* Sequence, linkage mapping and early developmental expression of the intestinal-type fatty acid-binding protein gene (*fabp2*) from zebrafish (*Danio rerio*). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **138**, 391–398 (2004).
- Grigoryev, D. N. *et al.* Cytochrome P450c17-expressing *Escherichia coli* as a first-step screening system for 17 α -hydroxylase-C_{17,20}-lyase inhibitors. *Anal. Biochem.* **267**, 319 (1999).
- Rheume, E. *et al.* Structure and expression of a new complementary DNA encoding the almost exclusive 3 beta-hydroxysteroid dehydrogenase/delta 5-delta 4-isomerase in human adrenals and gonads. *Mol. Endocrinol.* **5**, 1147–1157 (1991).
- Jesuthasan, S. & Stahle, U. Dynamic microtubules and specification of the zebrafish embryonic axis. *Curr. Biol.* **7**, 31–42 (1997).
- Solnica-Krezel, L. & Driever, W. Microtubule arrays of the zebrafish yolk cell: organization and function during epiboly. *Development* **120**, 2443–2455 (1994).
- Strahle, U. & Jesuthasan, S. Ultraviolet irradiation impairs epiboly in zebrafish embryos: evidence for a microtubule-dependent mechanism of epiboly. *Development* **119**, 909–919 (1993).
- Akhmanova, A. *et al.* Clasps are CLIP-115 and -170 associating proteins involved in the regional regulation of microtubule dynamics in motile fibroblasts. *Cell* **104**, 923–935 (2001).
- Liu, L., Tommasi, S., Lee, D. H., Dammann, R. & Pfeifer, G. P. Control of microtubule stability by the RASSF1A tumour suppressor. *Oncogene* **22**, 8125–8136 (2003).
- Dent, E. W. & Gertler, F. B. Cytoskeletal dynamics and transport in growth cone motility and axon guidance. *Neuron* **40**, 209–227 (2003).
- Watanabe, T., Noritake, J. & Kaibuchi, K. Regulation of microtubules in cell migration. *Trends Cell Biol.* **15**, 76–83 (2005).
- Morris, N. R. Nuclear positioning: the means is at the ends. *Curr. Opin. Cell Biol.* **15**, 54–59 (2003).
- Wittmann, T., Hyman, A. & Desai, A. The spindle: a dynamic assembly of microtubules and motors. *Nature Cell Biol.* **3**, E28–E34 (2001).
- Murakami, K., Fellous, A., Baulieu, E. E. & Robel, P. Pregnenolone binds to microtubule-associated protein 2 and stimulates microtubule assembly. *Proc. Natl. Acad. Sci. USA* **97**, 3579–3584 (2000).
- D'Amico, L. A. & Cooper, M. S. Morphogenetic domains in the yolk syncytial layer of axiating zebrafish embryos. *Dev. Dyn.* **222**, 611–624 (2001).
- Arensburg, J., Payne, A. H. & Orly, J. Expression of steroidogenic genes in maternal and extraembryonic cells during early pregnancy in mice. *Endocrinology* **140**, 5220–5232 (1999).
- Hu, M. C. *et al.* Steroid deficiency syndromes in mice with targeted disruption of *Cyp11a1*. *Mol. Endocrinol.* **16**, 1943–1950 (2002).
- Venihaki, M., Carrigan, A., Dikkes, P. & Majzoub, J. A. Circadian rise in maternal glucocorticoid prevents pulmonary dysplasia in fetal mice with adrenal insufficiency. *Proc. Natl. Acad. Sci. USA* **97**, 7336–7341 (2000).
- Kozlova, T. & Thummel, C. S. Essential roles for ecdysone signaling during *Drosophila* mid-embryonic development. *Science* **301**, 1911–1914 (2003).
- Cato, A. C., Nestl, A. & Mink, S. Rapid actions of steroid receptors in cellular signaling pathways. *Sci. STKE* **2002**, RE9 (2002).
- Schmidt, B. M. *et al.* Rapid, nongenomic steroid actions: A new age? *Front. Neuroendocrinol.* **21**, 57–94 (2000).
- Chiang, E. F. *et al.* Two *sox9* genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. *Dev. Biol.* **231**, 149–163 (2001).
- Gard, D. L. Organization, nucleation, and acetylation of microtubules in *Xenopus laevis* oocytes: a study by confocal immunofluorescence microscopy. *Dev. Biol.* **143**, 346–362 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank V. Korzh and C.-H. Hu for the plasmids for *in situ* hybridization, and K. Deen and M. Wyatt for editing the manuscript. This work was supported by grants from the National Science Council and Academia Sinica, Taiwan.

Author Contributions H.-J.H. performed all the experiments; M.-R.L. synthesized F-P₅; C.-T.C. devised and supervised the F-P₅ synthesis scheme; B.-c.C. oversaw the execution of the entire project.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.-c.C. (mbchung@sinica.edu.tw).

LETTERS

Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation

Mitsuhiro Watanabe^{1*}, Sander M. Houten^{1*}, Chikage Matak¹, Marcelo A. Christoffolete², Brian W. Kim², Hiroyuki Sato¹, Nadia Messaddeq¹, John W. Harney², Osamu Ezaki³, Tatsuhiko Kodama⁴, Kristina Schoonjans¹, Antonio C. Bianco² & Johan Auwerx^{1,5}

While bile acids (BAs) have long been known to be essential in dietary lipid absorption and cholesterol catabolism, in recent years an important role for BAs as signalling molecules has emerged. BAs activate mitogen-activated protein kinase pathways^{1,2}, are ligands for the G-protein-coupled receptor (GPCR) TGR5^{3,4} and activate nuclear hormone receptors such as farnesoid X receptor α (FXR- α ; NR1H4)^{5–7}. FXR- α regulates the enterohepatic recycling and biosynthesis of BAs by controlling the expression of genes such as the short heterodimer partner (SHP; NR0B2)^{8,9} that inhibits the activity of other nuclear receptors. The FXR- α -mediated SHP induction also underlies the downregulation of the hepatic fatty acid and triglyceride biosynthesis and very-low-density lipoprotein production mediated by sterol-regulatory-element-binding protein 1c¹⁰. This indicates that BAs might be able to function beyond the control of BA homeostasis as general metabolic integrators. Here we show that the administration of BAs to mice increases energy expenditure in brown adipose tissue, preventing obesity and resistance to insulin. This novel metabolic effect of BAs is critically dependent on induction of the cyclic-AMP-dependent thyroid hormone activating enzyme type 2 iodothyronine deiodinase (D2) because it is lost in D2^{–/–} mice. Treatment of brown adipocytes and human skeletal myocytes with BA increases D2 activity and oxygen consumption. These effects are independent of FXR- α , and instead are mediated by increased cAMP production that stems from the binding of BAs with the G-protein-coupled receptor TGR5. In both rodents and humans, the most thermogenically important tissues are specifically targeted by this mechanism because they coexpress D2 and TGR5. The BA–TGR5–cAMP–D2 signalling pathway is therefore a crucial mechanism for fine-tuning energy homeostasis that can be targeted to improve metabolic control.

To characterize the metabolic effects of BAs, we fed C57BL/6J mice on a high-fat (HF) diet supplemented with cholic acid (CA) at 0.5% w/w. The CA-containing diet was well tolerated and devoid of side effects for up to 47 days. Absence of toxicity was supported by the fact that feeding with CA did not affect food intake, nutrient uptake or serum liver enzymes¹¹ (Fig. 1a, and not shown). Whereas mice on the HF diet gained more weight than controls, supplementation of the HF diet with CA reduced weight gain comparably to that in chow-fed animals (Fig. 1b). Supplementation of chow with CA had no effect on gain in body weight (ref. 11 and not shown). The changes in weight gain reflect mainly modifications in adiposity: epididymal white adipose tissue (WAT), mesenteric WAT and intrascapular brown adipose tissue (BAT) of HF-fed animals were significantly increased

in weight. BAT of HF-fed animals was paler, indicative of decreased metabolic activity and increased fat accumulation. In addition, there was an expansion of WAT surrounding the BAT. Diet supplementation with CA completely prevented the HF-induced changes in adipose mass and morphology (Fig. 1c and Supplementary Fig. S1a). Diet supplementation with CA reversed 120 days of diet-induced weight gain (Fig. 1d). When obese mice were switched to an HF diet containing CA (F-FA), they normalized body weight within 30 days, an effect that was fully accounted for by a decrease in WAT mass (Fig. 1e). Besides preventing and reversing diet-induced obesity, CA also improved glucose tolerance in C57BL/6J mice with diet-induced obesity (ref. 11 and Supplementary Fig. S1b). In an independent experiment, C57BL/6J mice were fed on an HF diet supplemented with the synthetic FXR- α agonist GW4064, but no protective effect against diet-induced obesity was observed (Fig. 1f). The addition of CA to chow also decreased the obesity in KK-A^y mice, which was explained by a decrease in WAT mass. This decrease in body weight improved glucose tolerance (Supplementary Fig. S1c). The effects of CA feeding on energy homeostasis coincided with an increase in BA pool size and serum BA levels and a major change in BA composition (Fig. 1g). CA, tauroCA (TCA), deoxyCA (DCA) and taurodeoxyCA (TDCA) were markedly increased. These changes are likely to influence the different BA signalling pathways. These results show that a CA diet prevents and reverses fat accumulation and associated metabolic defects by means of a mechanism that does not depend solely on FXR- α .

With indirect calorimetry, higher CO₂ production and O₂ consumption were evident in animals fed on an HF diet containing CA when compared with animals on HF diet or chow (Fig. 2a), indicating increased energy expenditure. CA feeding attenuated the HF-induced decrease in the respiratory quotient (RQ), proving that CA increased fat oxidation. To identify the site(s) responsible for the increased energy expenditure, we performed a detailed histological analysis of key metabolic tissues. The HF diet caused adipocyte hypertrophy in WAT and BAT that was not observed when the HF diet was supplemented with CA (Supplementary Fig. S2a). Electron microscopic analysis of BAT demonstrated that, in comparison with the HF diet, CA supplementation increased the number of lamellar cristae in the mitochondria, indicating a major role for BAT in CA-induced energy expenditure (Fig. 2b).

We performed a microarray experiment comparing the expression profile in liver and BAT from mice fed on an HF diet or on an HF diet mixed with CA or chenodeoxycholic acid (CDCA). Major changes in hepatic gene expression were related to cholesterol and BA synthesis

¹Institut de Génétique et Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, 1 Rue Laurent Fries, 67404 Illkirch, France. ²Department of Medicine, Thyroid Section, Division of Endocrinology, Diabetes and Hypertension, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. ³Division of Clinical Nutrition, National Institute of Health and Nutrition, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8636, Japan. ⁴Laboratory for Systems Biology and Medicine, RCAST, University of Tokyo, Tokyo 153-8904, Japan. ⁵Institut Clinique de la Souris, 67404 Illkirch, France.

*These authors contributed equally to this work.

but not to energy homeostasis (not shown). In BAT, however, several genes involved in the control of energy expenditure were upregulated by BAs (Supplementary Table 1). We focused on *D2*, whose expression was most affected. *D2* converts inactive thyroxine (T_4) into active 3,5,3'-tri-iodothyronine (T_3)¹² and is a crucial determinant of thyroid hormone receptor saturation in cells in which it is expressed. The vital role of *D2* in energy homeostasis is well established in mice, in which BAT adaptive thermogenesis depends on this enzyme¹².

We used quantitative reverse transcriptase-mediated polymerase chain reaction to confirm the induction of *D2* by BAs and to measure messenger RNA levels of genes involved in energy homeostasis and BA signalling in liver, muscle and BAT. For these studies we used organs from the C57BL/6J mice described in Fig. 1a–c. The expression of key genes in energy expenditure such as the *peroxisome-proliferator-activated receptor γ coactivator (PGC)-1 α* , *PGC-1 β* , *uncoupling protein (UCP)-1*, *UCP-3*, *straight-chain acyl-CoA oxidase 1 (ACO)*, *muscle-type carnitine palmitoyltransferase I (mCPT-I)* and *D2* was significantly increased in BAT after BA feeding (Fig. 2c). Similar changes were not observed in liver or muscle. *FXR- α* and *SHP* expression was not detected in BAT under any of the conditions, whereas we confirmed the induction of *SHP* by BAs in liver (Fig. 2c)^{8,9}. *FGFR-4*, another gene implicated in BA signalling¹³, was only expressed to substantial amounts in liver. In addition, treatment with GW4064 did not induce a thermogenic gene response as seen with CA (not shown). These results confirm that BAT is the primary target of the metabolic effects of BAs in mice and that the beneficial effects of BAs on energy homeostasis do not depend solely on *FXR- α* .

Next we used mice with a targeted disruption of the *D2* gene (*D2*^{-/-} mice) and littermate controls in a diet-induced obesity study. Wild-type and *D2*^{-/-} mice gained weight to similar extents, indicating that adaptive thermogenesis was increased sufficiently in *D2*^{-/-} mice to prevent disproportional weight gain (Fig. 3a). This stands in contrast with the situation of acute exposure to cold, in which *D2*^{-/-} BAT thermogenesis is insufficient to maintain body temperature and

survival is possible only with a major increase in sympathetic activity¹⁴ and mechanical thermogenesis¹⁵. CA supplementation to the HF diet prevented diet-induced obesity in wild-type mice but had no effect in *D2*^{-/-} mice (Fig. 3a). Similarly, CA was unable to decrease the weight of the epididymal WAT and BAT in HF-fed *D2*^{-/-} mice (Fig. 3b). Histological and electron microscopic analysis of BAT showed that combining the HF diet with CA prevented lipid accumulation in BAT of wild-type but not that of *D2*^{-/-} mice (Fig. 3c and Supplementary Fig. S2b). The lack of alterations in serum T_4 and T_3 levels in mice fed CA (Supplementary Table 2) confirms the tissue specificity of our findings. Together, these results prove that *D2* is an essential mediator in the prevention of diet-induced weight gain by BAs.

Our data led us to focus on *FXR- α* -independent signalling events stimulated by BAs. *D2* expression is regulated primarily through the cAMP–protein kinase A (PKA) pathway, and a functional cAMP-responsive element (CRE) has been identified in human and mouse *D2* promoter¹². TGR5 is the only GPCR that has been reported to respond to BAs by the production of cAMP and the subsequent activation of PKA signalling pathways^{3,4}. We therefore added CA and DCA and their taurine conjugates (TCA and TDCA) to cells co-transfected with a CRE-driven luciferase reporter and a TGR5 expression plasmid. All tested BAs enhanced luciferase activity in a TGR5-dependent fashion, showing that cAMP production by BAs requires TGR5 (Fig. 3d). The effect of forskolin, an adenylyl cyclase activator, on luciferase activity was independent of TGR5 (Fig. 3d). Physiological concentrations of BAs increased luciferase expression dose dependently in TGR5-transfected cells, proving tight coupling between BA–TGR5 and intracellular PKA signalling (Fig. 3e). *D2* and *TGR5* mRNAs are coexpressed in mouse BAT, which has the highest relative expression level of both genes in the selected mouse tissues tested (Fig. 3f). Expression of *TGR5* and low levels of *D2* mRNAs were also found in brain and thyroid gland, but the latter has no measurable *D2* activity¹². Further proof that BAs signal by means of cAMP–PKA activation comes from elevated cAMP levels in BAT of C57BL/6J and KK-A^y mice fed on a CA diet (Fig. 3g). Taken together, these results indicate that the effects of BAs on BAT could be

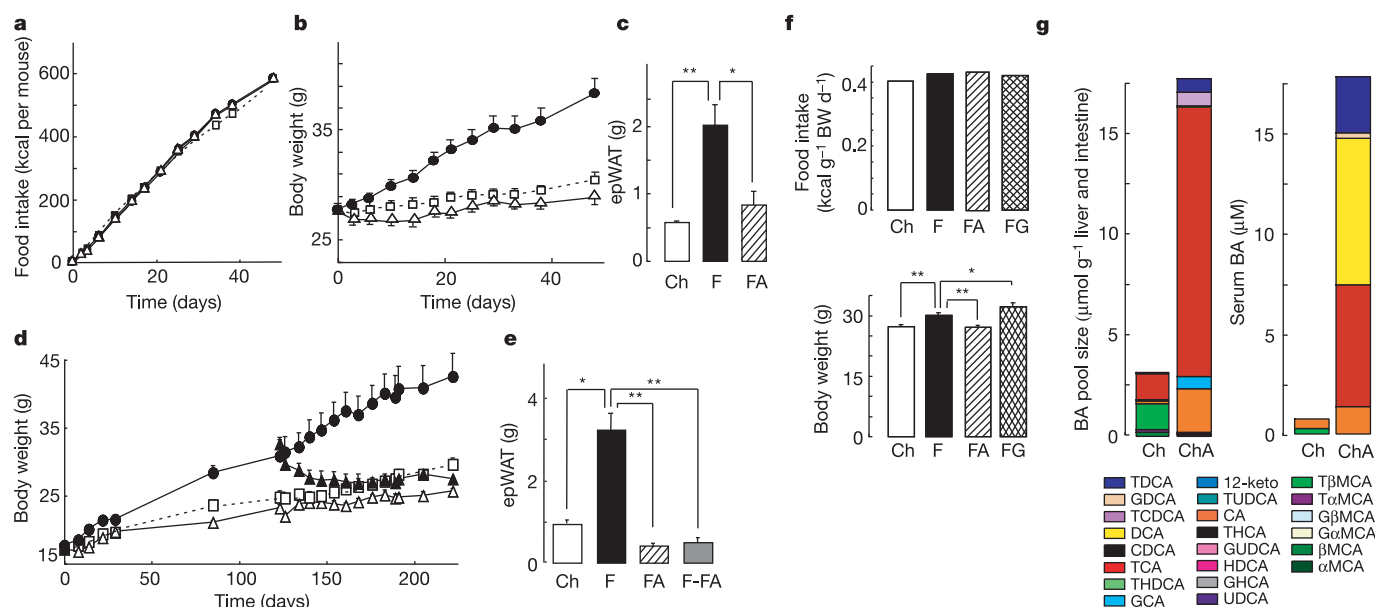


Figure 1 | CA decreases adiposity of mice fed a HF diet. **a, b**, Change in cumulative food intake (**a**) and body weight (**b**) of C57BL/6J mice over 47 days. Squares, chow (Ch); circles, HF diet (F); triangles, HF diet plus CA (FA). **c**, Comparison of epididymal WAT (epWAT). **d**, Changes of body weight in C57BL/6J mice. After 120 days half of the mice on the HF diet (filled triangles) were switched to HF diet supplemented with CA. Other symbols as in **a**. **e**, Comparison of epWAT weights. F-FA, switched to HF diet

supplemented with CA. **f**, Food intake and body weight (BW) of C57BL/6J mice after 1 month on diets containing natural CA or synthetic (GW4064) *FXR- α* agonist (FG, HF plus GW4064). **g**, Composition of BAs in enterohepatic organs and serum of KK-A^y mice after 21 days on the indicated diets. Abbreviations: Glyco (G), Hyo (H), Urso (U) and Muri (M). Error bars show s.e.m. ChA, chow diet supplemented with cholic acid.

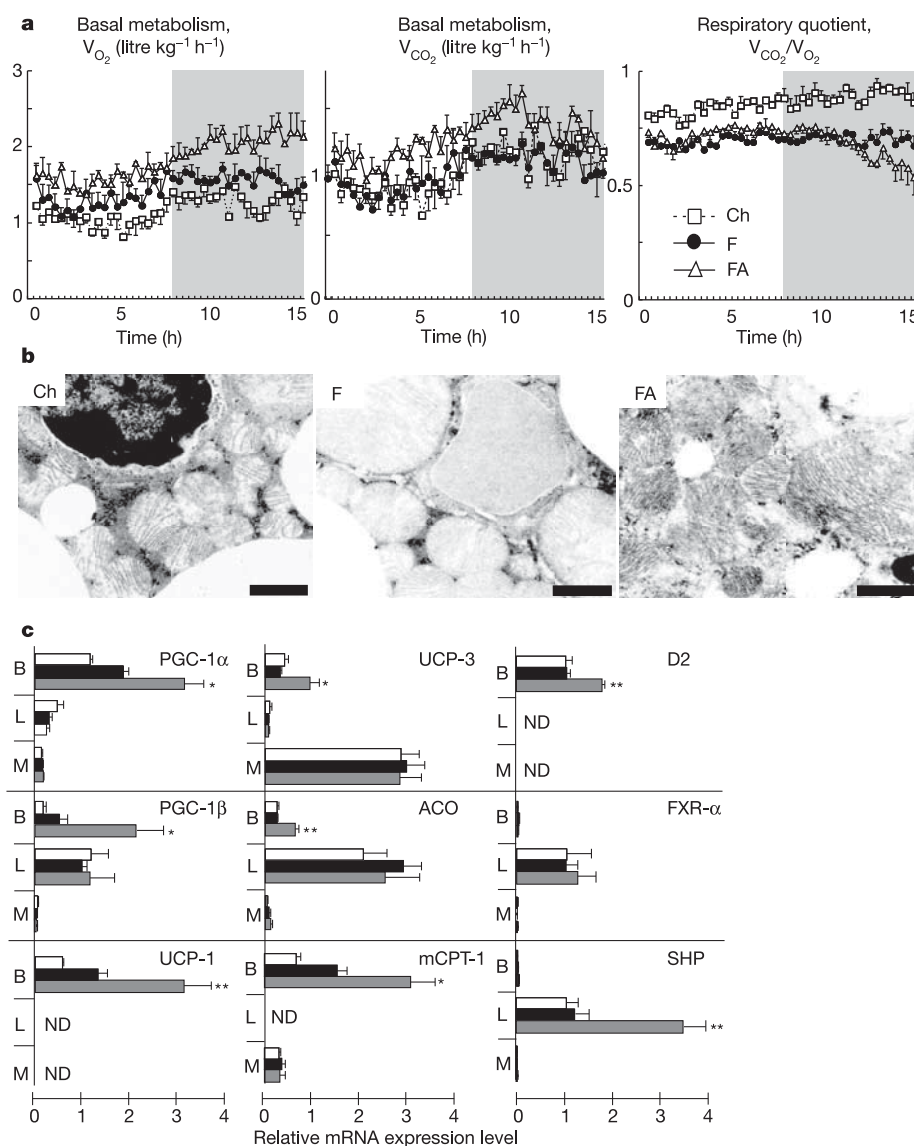


Figure 2 | CA increases energy expenditure. **a**, O_2 consumption, CO_2 production and RQ in mice on different diets for 4 months (C57BL/6J, $n = 3$, age 26 weeks). The acclimation time was 2 h. O_2 consumption was normalized to (body weight) $^{0.75}$. The shaded area indicates the dark phase. Squares, chow (Ch); circles, HF diet (F); triangles, HF diet plus CA (FA).

b, BAT analysis by transmission electron microscopy. Scale bar, 1 μm .

c, Relative mRNA expression levels of PGC-1 α , PGC-1 β , UCP-1, UCP-3, ACO, mCPT-1, D2, FXR- α and SHP in BAT (B), liver (L) and muscle (M). ND, not detectable. White bars, chow; black bars, HF diet; grey bars, HF diet plus CA. Error bars show s.e.m.

mediated through the activation of GPCRs, such as TGR5, and the subsequent activation of PKA signalling.

To reproduce the effect of BAs *in vitro*, we isolated BAT cells from C57BL/6J mice fed on either a control or an HF diet. TCA dose-dependently increased D2 mRNA and D2 activity (Fig. 4a). Interestingly, cells from HF-fed mice were more sensitive to the addition of TCA but also to forskolin (Fig. 4a). This could explain why BAs have major effects on energy homeostasis only when supplementing an HF diet and indicates that besides BAs other factors have a function in the induction of D2. Several different BAs increased cAMP, indicating that this second messenger induces D2 expression. The FXR- α agonist GW4064 failed to increase cAMP in BAT cells (Fig. 4b). The effect of an HF diet on cAMP production by BAT cells was less pronounced than on D2 mRNA and D2 activity, indicating that an HF diet influences D2 induction somewhere downstream of the GPCR.

In contrast to rodents, adult humans do not have significant amounts of BAT but express significant levels of D2 in skeletal muscle¹² (Fig. 4c), an organ of critical importance to energy homeo-

stasis. TGR5 mRNA was also expressed in human skeletal muscle (Fig. 4c). To investigate whether D2 expression in human skeletal muscle could be induced by BAs, by analogy with the induction of D2 expression in BAT in mice, we used primary cultures of human skeletal muscle myoblasts (HSMM) expressing D2 and high levels of TGR5 (Fig. 4c). On incubation of HSMM with TCA, we observed a significant and dose-dependent induction of D2 activity (Fig. 4d). Again, the FXR- α agonist GW4064 did not stimulate D2. Major BA species such as CA, TCA, DCA and CDCA increased cAMP levels in a dose-dependent fashion, paralleling the increase in D2 activity (Fig. 4e). However, GW4064 did not increase cAMP levels (Fig. 4e).

We were unable to knock down TGR5 function with several different short interfering RNAs as well as a plasmid encoding a small hairpin RNA. As an alternative approach to demonstrate that the effects of BAs on cAMP levels and D2 activity were mediated by TGR5, we used benzyl 2-keto-6-methyl-4-(2-thienyl)-1,2,3,4-tetrahydropyrimidine-5-carboxylate, a highly selective synthetic TGR5 agonist¹⁶ (Fig. 4f). This TGR5 agonist dose-dependently induced the

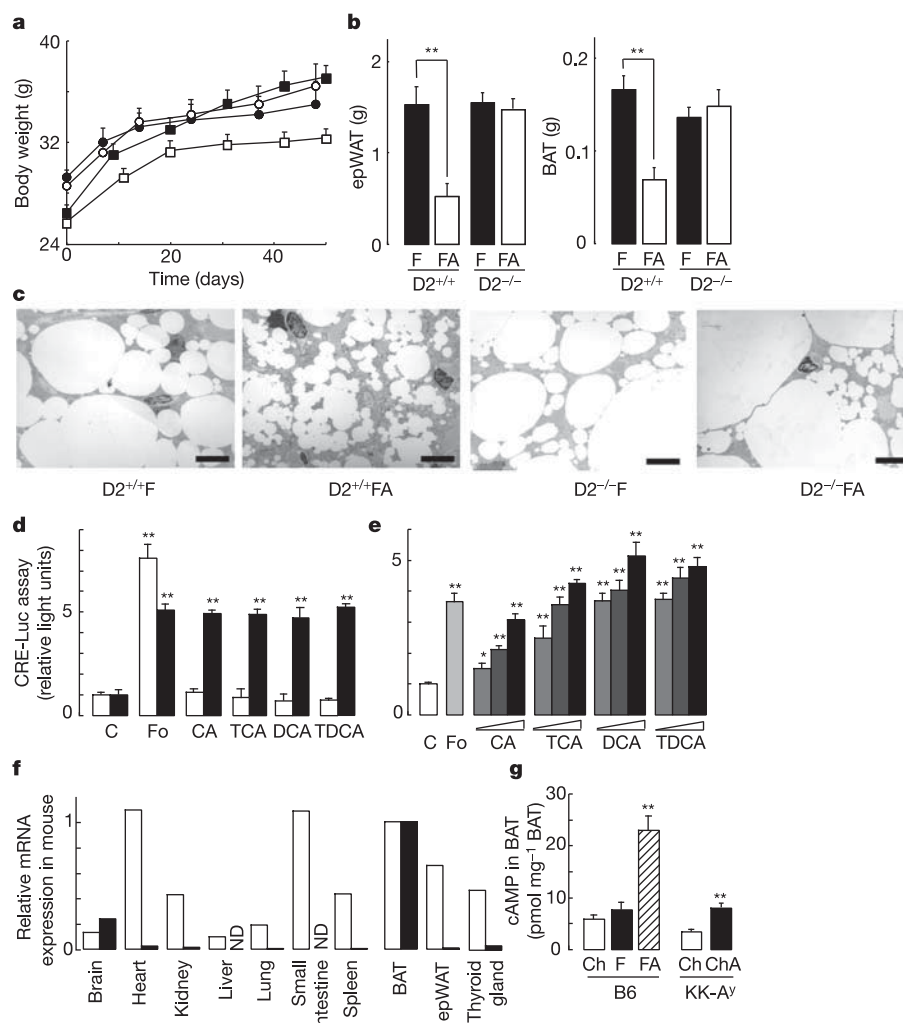


Figure 3 | BAs signal by means of D2. **a**, Body weight change in $D2^{+/+}$ and $D2^{-/-}$ mice over 50 days. Filled squares, $D2^{+/+}$, HF diet (F); open squares, $D2^{+/+}$, HF diet plus CA (FA); filled circles, $D2^{-/-}$, F; open circles, $D2^{-/-}$, FA. **b**, Comparison of the weights of epWAT and BAT. **c**, Osmium-tetroxide-stained BAT was analysed by transmission electron microscopy. Scale bars, 5 μ m. **d**, CRE reporter assay in CHO cells transfected with pCRE-Luc and TGR5 expression vector. Concentrations: 100 μ M BA, 5 μ M forskolin (Fo).

C means control. Open bars, vector; filled bars, pTGR5. **e**, CRE reporter assay in CHO cells transfected with pCRE-Luc and TGR5 expression vector in the presence of different concentrations (1.8, 5.5 and 17 μ M) of the indicated BAs. **f**, Expression of *TGR5* (open panels) and *D2* (filled panels) in selected mouse tissues. ND, not detectable. **g**, cAMP levels in BAT of C57BL/6J and KK-Ay mice after 7 days on the diets. Ch means chow and ChA means chow + CA. Error bars show s.e.m.

activity of a CRE-luciferase reporter in a TGR5-dependent fashion, but TCA was more effective at equimolar concentrations (Fig. 4f). In addition, the TGR5 agonist enhanced D2 activity in HSMM cells, an effect potentiated in the presence of 3-isobutyl-1-methylxanthine (Fig. 4g), indicating that the effects of BAs on D2 activity might be mediated by TGR5.

Finally, we determined the oxygen consumption and extracellular acidification rate in HSMM after treatment with BA. Oxygen consumption is a measure for aerobic mitochondrial oxidation, whereas extracellular acidification rate is a measure for glycolysis, lactate production and anaerobic metabolism. TCA, like T_3 , elevated cellular oxygen consumption in line with the induction of mitochondrial activity (Fig. 4h). This indicates that TCA induces D2, which converts T_4 derived from the fetal bovine serum into T_3 . In addition, we observed an increase in extracellular acidification rate for both T_3 and TCA after 72 h of treatment, which could suggest a limiting reducing capacity caused by oxygen shortage. In combination, these results suggest that BAs increase *D2* mRNA, *D2* activity and energy expenditure by means of a TGR5–cAMP-mediated pathway in mouse BAT and HSMM.

Recent observations indirectly support a role for D2 as a determinant of energy expenditure in human skeletal muscle. A functionally significant polymorphism in the *D2* gene¹⁷ is strongly associated with decreased whole-body glucose disposal rate, which is determined mainly by muscle glucose uptake¹⁸. In patients with type 2 diabetes this polymorphism is associated with increased insulin resistance¹⁷.

The pharmacological relevance of the BA–cAMP–D2– T_3 pathway is clearly supported by our data, but the same pathway could also be of physiological importance. The extraction of BAs from portal blood by the liver is remarkably efficient (70–90%) and usually remains constant during fasting and digestion¹⁹. Thus, after a meal, BA levels in the hepatocyte increase and a significant amount of BAs spills over into the systemic circulation. Fasting serum BA levels are usually lower than 5 μ M, whereas postprandial levels increase to up to 15 μ M (ref. 20), consistent with the concentrations required to stimulate cAMP production and D2 activity in our *in vitro* studies. Thus, the circadian variation in serum BA levels could be a diurnal hormonal signal reflecting food intake and one of the key factors in diet-induced thermogenesis.

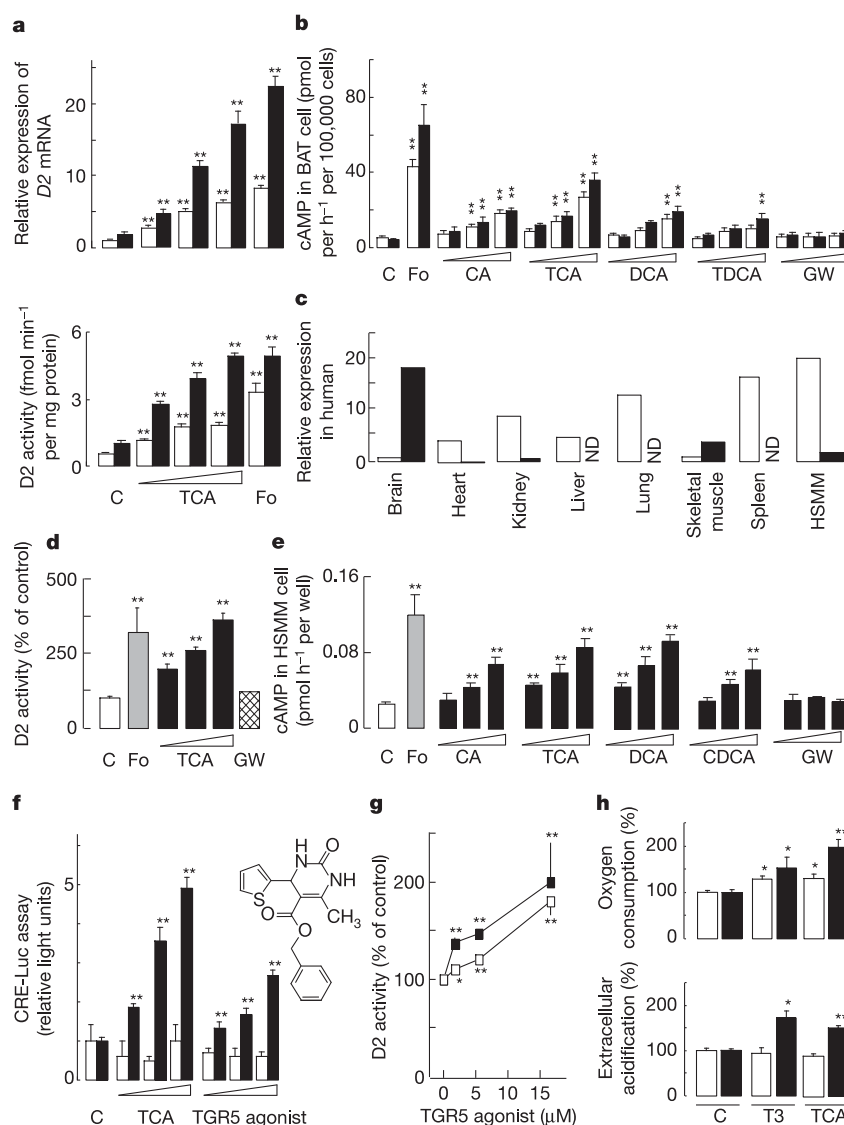


Figure 4 | *In vitro* effects of BAs. **a**, D2 expression (upper panel) and D2 activity (lower panel) in BAT cells from C57BL/6J mice after 14 days on chow (open columns) and HF diet (filled columns). Cells were treated with TCA or forskolin (Fo) (as in Fig. 3e). **b**, Induction of cAMP by BAs in BAT cells (as in Fig. 3e). GW, GW4064 at 1.1, 3.3 and 10 μ M. Open columns, chow; filled columns, HF diet. **c**, Expression of TGR5 (open columns) and D2 (filled columns) in selected human tissues. ND, not detectable. **d**, Induction of D2 activity in HSM cells by TCA (1.3, 4 and 12 μ M), GW4064 (3 μ M) and forskolin (10 μ M). **e**, Induction of cAMP by BAs in HSM cells (as in **b**). **f**, CRE reporter

assay (as in Fig. 3d). Agonists were used at 1, 3.2 and 10 μ M. Open columns, vector; filled columns, pTGR5. The structure of benzyl 2-keto-6-methyl-4-(2-thienyl)-1,2,3,4-tetrahydropyrimidine-5-carboxylate is also shown. **g**, Induction of D2 activity in HSM cells by the synthetic TGR5 agonist (1, 5 and 15 μ M) in the absence (open columns) or presence (filled columns) of 1 mM IBMX. **h**, Induction of oxygen consumption (upper panel) and extracellular acidification rate (lower panel) in HSM cells by 5 μ M TCA and 50 nM T₃. Open columns, 48 h; filled columns, 72 h. Error bars show s.e.m.

METHODS

Materials. 3-Isobutyl-1-methylxanthine (IBMX), CA, TCA, DCA, TDCA and forskolin were from Sigma. Benzyl 2-keto-6-methyl-4-(2-thienyl)-1,2,3,4-tetrahydropyrimidine-5-carboxylate was synthesized by ChemBridge. GW4064 was a gift from S. Inoue.

Plasmids. pCRE-Luc was from Clontech. NIH MGC Clone Collection MGC:40597 (pTGR5) was from Invitrogen.

Cell culture, transient transfection and luciferase assays. The Chinese hamster ovary cell line was obtained from ATCC. Cells were maintained in α -MEM with 10% (v/v) fetal bovine serum at 37 °C in a humidified atmosphere of 5% CO₂/95% air; 70–80% confluent cells were transfected in 96-well plates with Lipofectamine 2000 (Invitrogen). Each well contained 50 ng of luciferase reporter (pCRE-Luc) plasmid and 200 ng of TGR5 expression plasmid or the empty expression vector. After incubation for 6 h with DNA–Lipofectamine complexes, the transfection medium was exchanged for DMEM medium containing 0.1% BSA (Sigma). After incubation for 18 h, the medium was exchanged for 0.1% BSA–DMEM medium with BAs or forskolin. Stock

solutions of these compounds were prepared in dimethylsulphoxide. After incubation for 6 h with the different compounds, cells were lysed and used for luciferase determination.

mRNA expression analysis. mRNA expression levels were analysed in cDNA synthesized from total mRNA with the real-time polymerase chain reaction as described¹⁰. The sequences of the primer sets used are available at http://www-igbmc.u-strasbg.fr/recherche/Dep_GPSN/Eq_JAuwe/Publi/Paper.html.

Cell culture of HSM. HSM were obtained from Cambrex and were cultured in accordance with the supplier's instructions. D2 activity measurements were performed after incubation for 16 h with the indicated compounds. For cAMP measurements, cells were preincubated for 12 h in Opti-MEM (Life Technologies) supplemented with 1 mM IBMX. Then medium was changed and cells were incubated for a further 1 h in Opti-MEM with 1 mM IBMX, after which the medium was replaced by Opti-MEM with 1 mM IBMX and the indicated compounds. After 30 min, cells were lysed for cAMP analysis.

Oxygen consumption and extracellular acidification rate. HSM were seeded in 24-well microplates. Quadruplicate wells were treated with 50 nM T₃, 5 μ M

TCA or vehicle. Cells were assayed in the CellDoctor prototype instrument from Seahorse Bioscience to perform non-destructive, time-resolved measurements of oxygen consumption rate and extracellular acidification rate (ECAR). Immediately before measurement, medium was replaced with non-buffered pH 7.4 speciality medium (1:1 saline and 2 × DMEM) to facilitate rapid ECAR measurement. Three successive 10-min measurements were performed simultaneously at 5-min intervals in the quadruplicate wells. Immediately after measurement, total cell number was counted with a Beckman Coulter ViCell.

Cell isolation and cAMP generation in brown adipocytes. Freshly dispersed brown adipocytes were isolated as described²¹, with some modifications²². Cell integrity was confirmed by Trypan blue exclusion in the absence of BSA. Cells were counted and diluted to 500,000 cells ml⁻¹. Diluted cells were incubated for 1 h in the presence of 500 μM IBMX and 0.3 U ml⁻¹ adenosine deaminase. Responses to forskolin and BAs were generated as described with the experiment. All additions were diluted in the incubation medium (DMEM/F12). Incubation was terminated by the addition of perchloric acid to a final concentration of 6.0%.

Animals. Male C57BL/6J and KK-A^y mice, 6–7 weeks of age, were obtained from Charles River Laboratories France and CLEA Japan, respectively. D2^{-/-} mice were originally bred to a C57BL/6J-129SV background¹⁵ and backcrossed for five generations to C57BL/6J. The composition of the different diets was as described¹⁰. GW4064 was mixed with the HF diet at 180 mg kg⁻¹ of food. On the basis of their daily food intake, this resulted in a daily dose of 15 mg kg⁻¹ body weight. The mice were fasted for 4 h before blood and tissues were harvested for RNA isolation, lipid measurements and histology. Oxygen consumption was measured with the Oxymax apparatus (Columbus Instruments)²³. Histological analysis and transmission electron microscopy were performed as described²³.

Clinical biochemistry and evaluation of glucose, lipid and BA homeostasis. An oral glucose tolerance test was performed in animals that had been fasted overnight. Glucose was administered by gavage at a dose of 2 g kg⁻¹. An intraperitoneal insulin tolerance test was performed in animals fasted for 4 h. Insulin was injected at a dose of 0.75 U kg⁻¹ body weight. Glucose was quantified with the Maxi Kit Glucometer 4 (Bayer Diagnostic) or Glucose RTU (bioMérieux Inc.). Plasma insulin concentrations were measured by ELISA (Cristal Chem Inc.). D2 activity was assayed as described²⁴. cAMP was measured with the cAMP-Screen system (Applied Biosystems) or a solid-phase radioimmunoassay kit (NEN). BAs in plasma and enterohepatic organs of fed mice were determined as described in ref. 25.

Statistical analysis. Statistical differences were determined by either ANOVA or Student's *t*-test. Statistical significance is displayed as *P* < 0.05 (one asterisk) or *P* < 0.01 (two asterisks).

Received 11 April; accepted 19 October 2005.

Published online 8 January 2006.

- Gupta, S., Stravitz, R. T., Dent, P. & Hylemon, P. B. Down-regulation of cholesterol 7α-hydroxylase (CYP7A1) gene expression by bile acids in primary rat hepatocytes is mediated by the c-Jun N-terminal kinase pathway. *J. Biol. Chem.* **276**, 15816–15822 (2001).
- Qiao, L. *et al.* Bile acid regulation of C/EBPβ, CREB, and c-Jun function, via the extracellular signal-regulated kinase and c-Jun NH₂-terminal kinase pathways, modulates the apoptotic response of hepatocytes. *Mol. Cell. Biol.* **23**, 3052–3066 (2003).
- Kawamata, Y. *et al.* A G protein-coupled receptor responsive to bile acids. *J. Biol. Chem.* **278**, 9435–9440 (2003).
- Maruyama, T. *et al.* Identification of membrane-type receptor for bile acids (M-BAR). *Biochem. Biophys. Res. Commun.* **298**, 714–719 (2002).
- Makishima, M. *et al.* Identification of a nuclear receptor for bile acids. *Science* **284**, 1362–1365 (1999).
- Parks, D. J. *et al.* Bile acids: natural ligands for an orphan nuclear receptor. *Science* **284**, 1365–1368 (1999).
- Wang, H., Chen, J., Hollister, K., Sowers, L. C. & Forman, B. M. Endogenous bile acids are ligands for the nuclear receptor FXR/BAR. *Mol. Cell* **3**, 543–553 (1999).
- Goodwin, B. *et al.* A regulatory cascade of the nuclear receptors FXR, SHP-1, and LRH-1 represses bile acid biosynthesis. *Mol. Cell* **6**, 517–526 (2000).
- Lu, T. T. *et al.* Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. *Mol. Cell* **6**, 507–515 (2000).
- Watanabe, M. *et al.* Bile acids lower triglyceride levels via a pathway involving FXR, SHP, and SREBP-1c. *J. Clin. Invest.* **113**, 1408–1418 (2004).
- Ikemoto, S. *et al.* Cholate inhibits high-fat diet-induced hyperglycemia and obesity with acyl-CoA synthetase mRNA decrease. *Am. J. Physiol.* **273**, 37–45 (1997).
- Bianco, A. C., Salvatore, D., Gereben, B., Berry, M. J. & Larsen, P. R. Biochemistry, cellular and molecular biology, and physiological roles of the iodothyronine selenodeiodinases. *Endocr. Rev.* **23**, 38–89 (2002).
- Holt, J. A. *et al.* Definition of a novel growth factor-dependent signal cascade for the suppression of bile acid biosynthesis. *Genes Dev.* **17**, 1581–1591 (2003).
- Christoffolete, M. A. *et al.* Mice with targeted disruption of the Dio2 gene have cold-induced overexpression of the uncoupling protein 1 gene but fail to increase brown adipose tissue lipogenesis and adaptive thermogenesis. *Diabetes* **53**, 577–584 (2004).
- de Jesus, L. A. *et al.* The type 2 iodothyronine deiodinase is essential for adaptive thermogenesis in brown adipose tissue. *J. Clin. Invest.* **108**, 1379–1385 (2001).
- Hinuma, S. *et al.* Screening method (method for the screening of agonists or antagonists of TGR5). Japanese patent 2003–380574 (P2005–21151A (JP)). 2005.
- Canani, L. H. *et al.* The type 2 deiodinase A/G (Thr92Ala) polymorphism is associated with decreased enzyme velocity and increased insulin resistance in patients with type 2 diabetes mellitus. *J. Clin. Endocrinol. Metab.* **90**, 3472–3478 (2005).
- Mentuccia, D. *et al.* Association between a novel variant of the human type 2 deiodinase gene Thr92Ala and insulin resistance: evidence of interaction with the Trp64Arg variant of the beta-3-adrenergic receptor. *Diabetes* **51**, 880–883 (2002).
- Cohen, D. E. in *Hepatology: A Textbook of Liver Disease* (eds Zakim, D. & Boyer, T. D.) 1713–1743 (Saunders, Philadelphia, 2003).
- Everson, G. T. Steady-state kinetics of serum bile acids in healthy human subjects: single and dual isotope techniques using stable isotopes and mass spectrometry. *J. Lipid Res.* **28**, 238–252 (1987).
- Fain, J. N., Reed, N. & Saperstein, R. The isolation and metabolism of brown fat cells. *J. Biol. Chem.* **242**, 1887–1894 (1967).
- Branco, M., Ribeiro, M., Negrao, N. & Bianco, A. C. 3,5,3'-Triiodothyronine actively stimulates UCP in brown fat under minimal sympathetic activity. *Am. J. Physiol.* **276**, E179–E187 (1999).
- Picard, F. *et al.* SRC-1 and TIF2 control energy balance between white and brown adipose tissues. *Cell* **111**, 931–941 (2002).
- Curcio, C. *et al.* The human type 2 iodothyronine deiodinase is a selenoprotein highly expressed in a mesothelioma cell line. *J. Biol. Chem.* **276**, 30183–30187 (2001).
- Sakakura, H. *et al.* Simultaneous determination of bile acids in rat bile and serum by high-performance liquid chromatography. *J. Chromatogr.* **621**, 123–131 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Iwasaki, A. Izumi, T. Taniguchi, K. Sakai, G. Tsujimoto, Y. Kawamata, H. Overmars, T. Sorg, M.-F. Champy and the staff of the Institut Clinique de la Souris for technical assistance and discussions. We also thank Seahorse Bioscience for the collaborative studies of oxygen consumption and extracellular acidification rate in the human skeletal myocytes. Work in the laboratories of the authors is supported by grants from CNRS, INSERM, ULP, FRM, the Hôpital Universitaire de Strasbourg, the NIH, EMBO and the EU.

Author Contributions M.W. and S.M.H. were involved in project planning, experimental work and data analysis; C.M., M.A.C., B.W.K., H.S., N.M., J.W.H., O.E., T.K. and K.S. performed experimental work; and A.C.B. and J.A. were involved in project planning and data analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A. (auwerx@igbmc.u-strasbg.fr).

LETTERS

Architecture of ribonucleoprotein complexes in influenza A virus particles

Takeshi Noda^{1,2,3}, Hiroshi Sagara⁴, Albert Yen⁵, Ayato Takada^{3,6†}, Hiroshi Kida², R. Holland Cheng^{5,7} & Yoshihiro Kawaoka^{1,3,6,8}

In viruses, as in eukaryotes, elaborate mechanisms have evolved to protect the genome and to ensure its timely replication and reliable transmission to progeny. Influenza A viruses are enveloped, spherical or filamentous structures, ranging from 80 to 120 nm in diameter¹. Inside each envelope is a viral genome consisting of eight single-stranded negative-sense RNA segments of 890 to 2,341 nucleotides each¹. These segments are associated with nucleoprotein and three polymerase subunits, designated PA, PB1 and PB2; the resultant ribonucleoprotein complexes (RNPs) resemble a twisted rod (10–15 nm in width and 30–120 nm in length) that is folded back and coiled on itself^{2–4}. Late in viral infection, newly synthesized RNPs are transported from the nucleus to the plasma membrane, where they are incorporated into progeny virions capable of infecting other cells. Here we show, by transmission electron microscopy of serially sectioned virions, that the RNPs of influenza A virus are organized in a distinct pattern (seven segments of different lengths surrounding a central segment). The individual RNPs are suspended from the interior of the viral envelope at the distal end of the budding virion and are oriented perpendicular to the budding tip. This finding argues against random incorporation of RNPs into virions⁵, supporting instead a model in which each segment contains specific incorporation signals that enable the RNPs to be recruited and packaged as a complete set^{6–12}. A selective mechanism of RNP incorporation into virions and the unique organization of the eight RNP segments may be crucial to maintaining the integrity of the viral genome during repeated cycles of replication.

To elucidate the architecture of the virion interior, we longitudinally and transversely sectioned A/WSN/33 (H1N1) virions budding from Madin–Darby canine kidney (MDCK) cells at 10 h after infection. Although A/WSN/33 virions released into culture medium are spherical in shape¹³, the budding virions in longitudinal sections were elongated and contained rod-like structures that were associated with host-derived lipid bilayer envelopes and oriented perpendicular to the budding tip of the virion (Fig. 1a and Supplementary Fig. 1). They were ~12 nm in width and up to 130 nm in length, consistent with the sizes of purified RNPs^{2–4}. In transversely sectioned budding virions, electron-dense dots, representing the rod-like structures seen in longitudinal sections, were apparent inside each virion (Fig. 1b). Notably, many of the virions contained eight dots, arranged as seven in a circle surrounding one at the centre (Fig. 1c); not more than eight dots were observed in a virion. To determine the lengths of the rod-like structures, we compared serial transverse sections of whole budding virions (Fig. 1d and

Supplementary Fig. 2). In all virions examined, the number of dots decreased progressively with increasing distance from the budding tip of the virion. Together, these results indicate that influenza A virions contain a highly organized set of eight rod-like structures of different lengths (Fig. 2).

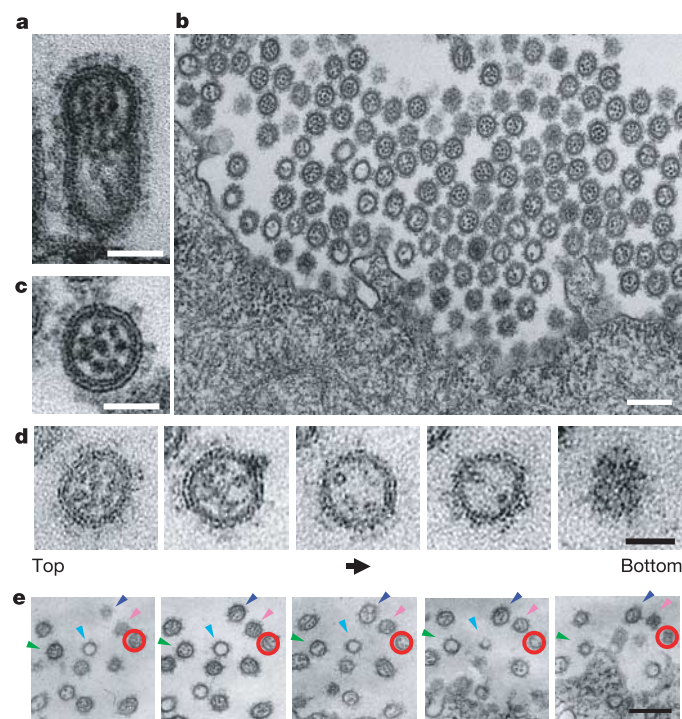


Figure 1 | Budding virions show a specific arrangement of eight rod-like structures of different lengths. **a**, Rod-like structures, 12 nm in width, are associated with the viral envelope at the distal end of the budding virion. **b**, **c**, Electron-dense dots ($N \leq 8$), representing transversely sectioned rods, were observed in each virion in a characteristic configuration (several peripheral dots surrounding a core dot). **d**, Serial sections of a virion cut from the distal end. As the distance from the end increased the number of dots decreased, suggesting that the eight rod-like structures differed in length. **e**, Lower magnification views of the serial ultrathin sections in **d**, demonstrating that the serial section shown in **d** represents the same virion (circled in red). Other virions are indicated by arrowheads. Scale bars, 50 nm (**a**, **c**, **d**); 200 nm (**b**, **e**).

¹Internal Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ²Laboratory of Microbiology, Department of Disease Control, Graduate School of Veterinary Medicine, Hokkaido University, Kita-ku, Sapporo 060-0818, Japan. ³Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. ⁴Fine Morphology Laboratory, Department of Basic Medical Science, and Division of Virology, Institute of Medical Science, University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁵Department of Biosciences, Karolinska Institute, 141 57 Huddinge, Sweden. ⁶Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁷Molecular and Cellular Biology, University of California, Davis, California 95616, USA. ⁸Department of Pathological Science, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. [†]Present address: Research Center for Zoonosis Control, Hokkaido University, Kita-ku, Sapporo 060-0818, Japan.

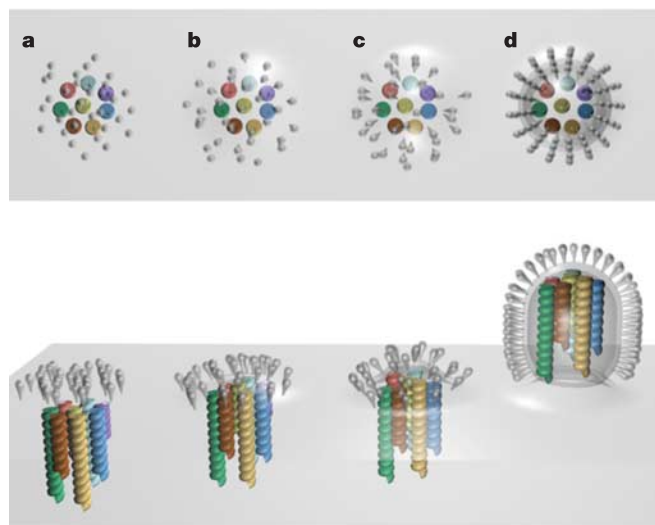


Figure 2 | Rod-like structures in a developing virion. **a–c**, Rod-like structures align vertically with the plasma membrane. **d**, The fully developed budding virion contains a set of eight rod-like structures. Virions in the upper row are viewed from the top; those in the lower row are shown in side profile.

To determine whether or not the arrangement of the virion interior in spherical A/WSN/33 virus is unique to that strain, we examined MDCK cells infected with other strains of spherical influenza A viruses isolated from humans (A/Aichi/2/68 (H3N2) and A/Puerto Rico/8/34 (H1N1)), ducks (A/duck/Hong Kong/836/80 (H3N1) and A/mallard/New York/6750/78 (H2N2)), and swine (A/swine/Chiba/1/91 (H1N2); see Supplementary Fig. 3). In all strains tested, the virions showed the same organization of rod-like structures, in which a central dot was surrounded by seven dots, duplicating the pattern seen in A/WSN/33 virions.

Because some influenza A viruses, especially newly isolated strains, are filamentous¹, we next examined the architecture of the filamentous A/Udorn/307/72 (H3N2) virus. Although most of the transverse sections of filamentous virions lacked electron-dense material (Supplementary Fig. 4a), some showed the typical configuration of seven dots with a single core dot (Supplementary Fig. 4b). In longitudinally sectioned filamentous virions, the rod-like structures were confined to the distal end of each filamentous particle (Supplementary Fig. 4c); the remainder of the virion was empty, consistent with the apparent

lack of internal structures in transverse sections of filamentous virions. Thus, both spherical and filamentous influenza virions budding from cells possess an organized set of eight rod-like structures that are associated with the envelope at the budding end of the virion.

Our serendipitous observation that some influenza A virions are partially disrupted when freeze-dried enabled us to study the nature of the rod-like structures inside virions. The widths of the structures released by freeze-dried, negatively stained A/Puerto Rico/8/34 virions were ~12 nm on average (Fig. 3a), similar to the widths of structures inside sectioned virions (Fig. 1). These released structures were morphologically indistinguishable from the RNPs purified from virions in previous studies^{2–4} and reacted with anti-nucleoprotein monoclonal antibodies conjugated to colloidal gold (Fig. 3b). To determine further the nature of the rod-like structures observed in virions, we performed ultrathin-section immunoelectron microscopy using mouse monoclonal antibody against nucleoprotein. The electron-dense dots in transversely sectioned virions reacted with anti-nucleoprotein monoclonal antibodies conjugated to gold particles (Fig. 3c), indicating that they were indeed viral RNPs.

To analyse the interior architecture of virions in greater detail, we used electron tomography, which constructs a three-dimensional density map of molecules, the 'tomogram', at a near-molecular resolution¹⁴. Taking computational thin slices from the map, Fig. 4 shows the results of electron tomography applied to virions budding from MDCK cells. Within a virion, outlined in Fig. 4 by a bilayered membrane with protruding viral glycoproteins, were eight RNPs with a distorted round or square shape. Although some of the RNPs were clearly isolated from each other, there were instances in which peripheral RNPs, as well as central and peripheral RNPs, appeared to be in close contact (Fig. 4, arrows). Whether such contact represents nonspecific association of the RNPs inside the virion or specific interactions with functional significance remains to be determined.

Our findings address a long-standing controversy in influenza virus research: are viral RNA segments incorporated randomly⁵ or selectively^{6,7} into virions? Together with studies showing that all viral RNPs possess segment-specific packaging signals (refs 8–12; and Y. Muramoto and Y.K., and M. Ozawa and Y.K., unpublished data), our morphological findings clearly favour a model of selective incorporation. Still unclear is how the eight RNPs are organized into a characteristic arrangement within virions. One possibility is that coding regions of the viral RNAs possess signal sequences that promote recruitment of the segments during virion assembly⁸, which may then control intersegmental association. Our study contributes

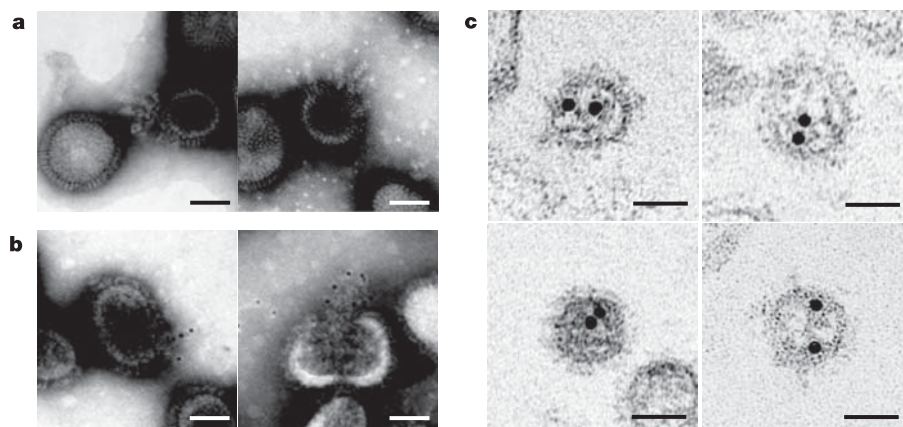


Figure 3 | Identification of rod-like structures in virions as viral RNPs. **a**, Virions ruptured by freeze-drying release twisted rods, 12 nm in width, which are indistinguishable from described purified RNPs^{2–4}. **b**, Twisted rods specifically immunolabelled with anti-nucleoprotein monoclonal antibodies

conjugated to 5-nm gold particles. **c**, Immunogold labelling of electron-dense dots within transversely sectioned virions with anti-NP monoclonal antibodies conjugated to 10-nm gold particles. Scale bars, 50 nm.

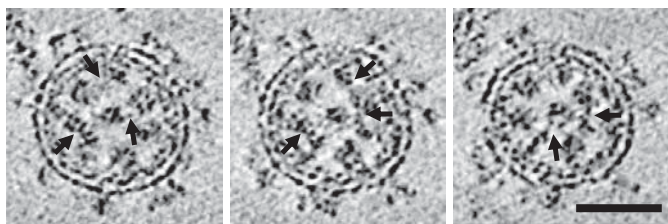


Figure 4 | Electron tomography of RNP complexes in a budding virion. The virion was reconstructed in three dimensions by electron tomography, and three tomograms of its upper portion (at 2.15-nm intervals) containing eight RNPs are shown. Note the close proximities (possible associations) of the RNPs (arrows). Scale bar, 50 nm.

fundamental knowledge to attempts aimed at elucidating the mechanisms of genome incorporation into virions. Defining the interior architecture of influenza virions may speed up the development of both antiviral compounds and more efficient methods of gene delivery and expression.

METHODS

Cells and viruses. A stock of A/WSN/33 (H1N1) was prepared by growing the virus in MDCK cells. Stocks of A/Aichi/2/68 (H3N2), A/Puerto Rico/8/34 (H1N1), A/duck/Hong Kong/836/80 (H3N1), A/mallard/New York/6750/78 (H2N2), A/swine/Chiba/1/91 (H1N2) and A/Udorn/307/72 (H3N2) were prepared in embryonated chicken eggs aged 10–11 d.

Ultrathin section electron microscopy. Electron microscopy was done as described¹⁵. In brief, MDCK cells were infected with virus at a multiplicity of infection of more than 10 and were then prefixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer (pH 7.4) on ice for 1 h. After being washed with the same buffer, the cells were post-fixed with 2% osmium tetroxide on ice for 1 h, stained with uranyl aqueous solution *en bloc*, dehydrated with a series of ethanol gradients followed by propylene oxide treatment, and embedded in Epon 812 Resin mixture. Ultrathin sections (~60-nm thick) were stained with 2% uranyl acetate in 70% ethanol for 3 min at room temperature and in Reynold's lead for 3 min at room temperature, and then examined with a JEM-1200EX electron microscope (Jeol) operated at 80 kV. For ultrathin section-immuno-electron microscopy, the sections were prepared on nickel grids as described above and incubated with saturated sodium periodate solution^{16,17}, followed by 0.2 M glycine in PBS buffer. After being washed with PBS, the sections were incubated with 1% bovine serum albumin (BSA) in PBS, and then with anti-influenza A virus nucleoprotein mouse monoclonal antibodies. They were then washed with PBS and incubated with a goat anti-mouse immunoglobulin conjugated to 10-nm gold particles (1:100 dilution, BBI International).

Preparation of serial ultrathin sections. Samples for serial ultrathin sections were prepared as described above. Sections (~20-nm thick) were continuously cut with an Ultracut S ultramicrotome (Leica). The resultant series of ultrathin sections were placed on a single-slot copper grid and examined with an H-7500 electron microscope (Hitachi) operated at 60 kV.

Negative staining of freeze-dried samples. Purified virions were adsorbed to Formvar-carbon-coated copper grids and negatively stained with 2% phosphotungstic acid solution (PTA). Grids were immediately placed in liquid nitrogen and transferred to a JEE-4X vacuum evaporator (Jeol) for freeze-drying. Immunoelectron microscopy was done with freeze-dried samples on nickel grids treated as described above, except that the 2% PTA staining step was omitted. The grids were then treated with 1% BSA in PBS and mouse monoclonal antibodies against influenza A virus nucleoprotein, washed with PBS, and incubated with a goat anti-mouse immunoglobulin conjugated to 5-nm gold particles (1:50 dilution, BBI International). After being washed, the samples were fixed with 2% glutaraldehyde and negatively stained with 2% PTA.

Tomographic reconstruction. Sections (~50-nm thick) were prepared as described above and affixed to 10-nm colloidal gold particles on the upper surface, which served as fiducial markers for subsequent image alignment. They were then placed in a CompStage specimen holder (FEI) and imaged in a Tecnai G2 Spha (FEI) operating at 200 kV. Images were taken at a magnification of

58,000 every 2° (from -65° to +70°) and recorded by a Gatan CCD camera (1,024 × 1,024 pixels with 0.43 nm per pixel) with an accumulative electron dose of 400 e⁻ Å⁻². The alignment of the projections was calculated by using IMOD software¹⁸ and the 10-nm gold beads, and the three-dimensional reconstruction was based on an *R*-weighted back projection. Galleries of slice tomograms were displayed with Amira software (Template Graphics) at intervals of 2.15-nm voxels.

Received 26 August; accepted 26 October 2005.

- Lamb, R. A. & Krug, R. M. in *Fields Virology* (eds Knipe, D. M. & Howley, P. M.) 1487–1532 (Lippincott, Williams & Wilkins, Philadelphia, 2001).
- Compans, R. W., Content, J. & Duesberg, P. H. Structure of ribonucleoprotein of influenza virus. *J. Virol.* **10**, 795–800 (1972).
- Heggeness, M. H. *et al.* Studies on the helical nucleocapsid of influenza virus. *Virology* **118**, 466–470 (1982).
- Oxford, J. S. & Hockley, D. J. *Orthomyxoviridae. Animal Virus Structure* 213–232 (Elsevier, Amsterdam, 1987).
- Enami, M., Sharma, G., Benham, C. & Palese, P. An influenza virus containing nine different RNA segments. *Virology* **185**, 291–298 (1991).
- Duhaut, S. D. & McCauley, J. W. Defective RNAs inhibit the assembly of influenza virus genome segments in a segment-specific manner. *Virology* **216**, 326–337 (1996).
- Odagiri, T. & Tashiro, M. Segment-specific noncoding sequences of the influenza virus genome RNA are involved in the specific competition between defective interfering RNA and its progenitor RNA segment at the virion assembly step. *J. Virol.* **71**, 2138–2145 (1997).
- Fujii, Y. *et al.* Selective incorporation of influenza virus RNA segment into virions. *Proc. Natl Acad. Sci. USA* **100**, 2002–2007 (2003).
- Watanabe, T. *et al.* Exploitation of nucleic acid packaging signals to generate a novel influenza virus-based vector stably expressing two foreign genes. *J. Virol.* **77**, 10575–10583 (2003).
- Fujii, K. *et al.* Importance of both the coding and the segment-specific noncoding regions of the influenza A virus NS segment for its efficient incorporation into virions. *J. Virol.* **79**, 3766–3774 (2005).
- Liang, Y., Hong, Y. & Parslow, T. G. *cis*-Acting packaging signals in the influenza virus PB1, PB2, and PA genomic RNA segments. *J. Virol.* **79**, 10348–10355 (2005).
- Dos Santos Afonso, E. *et al.* The generation of recombinant influenza A viruses expressing a PB2 fusion protein requires the conservation of a packaging signal overlapping the coding and noncoding regions at the 5' end of the PB2 segment. *Virology* **341**, 34–46 (2005).
- Schulze, I. T. The structure of influenza virus. *Virology* **42**, 890–904 (1970).
- Medalia, O. *et al.* Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science* **298**, 1209–1213 (2002).
- Noda, T. *et al.* Ebola virus VP40 drives the formation of virus-like filamentous particles along with GP. *J. Virol.* **76**, 4855–4865 (2002).
- Bendayan, M. & Zollinger, M. Ultrastructural localization of antigenic sites on osmium-fixed tissues applying the protein A–gold technique. *J. Histochem. Cytochem.* **31**, 101–109 (1983).
- Bendayan, M. & Maestracci, N. D. Pituitary adenomas: patterns of hPRL and hGH secretion as revealed by high resolution immunocytochemistry. *Biol. Cell* **52**, 129–138 (1984).
- Kremer, J. R., Mastrorade, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Gilbert for editing the manuscript; M. Imai, Y. Muramoto and K. Fujii for discussion; Y. Hirata, K. Aoyama and K. Inoke for technical assistance with electron microscopic tomography; and Y. Kawaoka for illustrations. This work was supported by CREST (Japan Science and Technology Agency), by Grants-in-Aid by the Ministry of Education, Culture, Sports, Science and Technology, by the Ministry of Health, Labor and Welfare, Japan, and by a National Institute of Allergy and Infectious Disease Public Health Service research grant (to Y.K.); and by Swedish Research Council grants and the STINT Foundation (to R.H.C.). T.N. was the recipient of a fellowship from the incorporated foundation SYOUSHISYA and a research fellowship from the Japan Society for the Promotion of Science for Young Scientists.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.K. (Kawaoka@ims.u-tokyo.ac.jp).

Chromatin organization and cell fate switch respond to positional information in *Arabidopsis*

Silvia Costa¹ & Peter Shaw¹

Many types of plant cell retain their developmental plasticity and have the capacity to switch fate when exposed to a new source of positional information. In the root epidermis of *Arabidopsis*, cells differentiate in alternating files of hair cells and non-hair cells^{1,2}, in response to positional information and the activity of the homeodomain transcription factor *GLABRA2* (*GL2*) in future non-hair cells^{3–6}. Here we show by three-dimensional fluorescence *in situ* hybridization on intact root epidermal tissue that alternative states of chromatin organization around the *GL2* locus are required to control position-dependent cell-type specification. When, as a result of an atypical cell division, a cell is displaced from a hair file into a non-hair file, it switches fate⁶. We show that during this event the chromatin state around the *GL2* locus is not inherited, but is reorganized in the G1 phase of the cell cycle in response to local positional information. This ability to remodel chromatin organization may provide the basis for the plasticity in plant cell fate changes.

In the meristematic epidermis of the *Arabidopsis* root, trichoblasts and atrichoblasts divide transversely to form alternating files of hair and non-hair cells along the longitudinal axis of the root^{1,2} (Supplementary Fig. 1). This patterned organization of the epidermis is established during embryogenesis and is then maintained post-embryonically in the seedling root^{6–8}, indicating a mechanism that propagates the epidermal pattern during root development. Analysis of epidermal T clones shows, however, that in the seedling root epidermal cell fate is not specified by lineage but by position, whereby atrichoblasts overlay a single cortical cell, and trichoblasts overlay the cleft between two cortical cells⁶ (Supplementary Fig. 1). Epidermal T clones derive from the occasional division of an epidermal cell along the longitudinal plane that gives rise to two daughter cells, which in turn will form two new parallel files with distinct fates, depending on their position relative to the underlying cortical cells (Supplementary Fig. 1). Characterization of the *scrambled* (*scm*) mutant, which is impaired in the activity of a receptor-like kinase, has confirmed the importance of positional information and its interpretation in guiding root epidermis specification⁹. Here we sought to elucidate the role of chromatin organization in position-dependent cell-type specification.

First, we determined whether alternative chromatin states are associated with different cell types by analysing the genomic region around the *GL2* locus. We hybridized, by three-dimensional fluorescence *in situ* hybridization (3D FISH), intact meristematic epidermal tissue from wild-type roots using the *GL2*BAC probe (Methods). In nuclei of atrichoblasts, where *GL2* is expressed^{3–6}, bright FISH signals were detected, indicating that chromatin is easily accessible to this probe in an 'open' conformation. By contrast, in nuclei of trichoblasts, where *GL2* is not expressed^{3–6}, no signals or only very faint signals were detected, indicating that this genomic region is sequestered from interaction with the probe in a 'closed' conformation (Fig. 1a, b, and Supplementary Fig. 2a, b). This result shows that the

chromatin state around the *GL2* locus differs in the two types of cell, linking chromatin state, expression of *GL2* and cell specification in root epidermal cells.

To determine whether or not *GL2* expression itself is directly responsible for the alternative chromatin organization at the locus, we analysed *cpc* and *wer* mutants. *WEREWOLF* (*WER*) and *CAPRICE* (*CPC*) encode MYB proteins required to establish the patterned organization of the root epidermis by controlling the position-dependent expression of *GL2* (refs 10–13). In *cpc*, *GL2* is expressed in every epidermal cell and all epidermal cells differentiate as non-hair cells^{11–13}. In *wer*, *GL2* and *CPC* are not expressed and all epidermal cells acquire a hair fate^{10,11}. We hybridized *cpc* and *wer* meristematic epidermal tissue by FISH using the *GL2*BAC probe. In both *cpc* and *wer*, we detected strong but equivalent FISH signals in

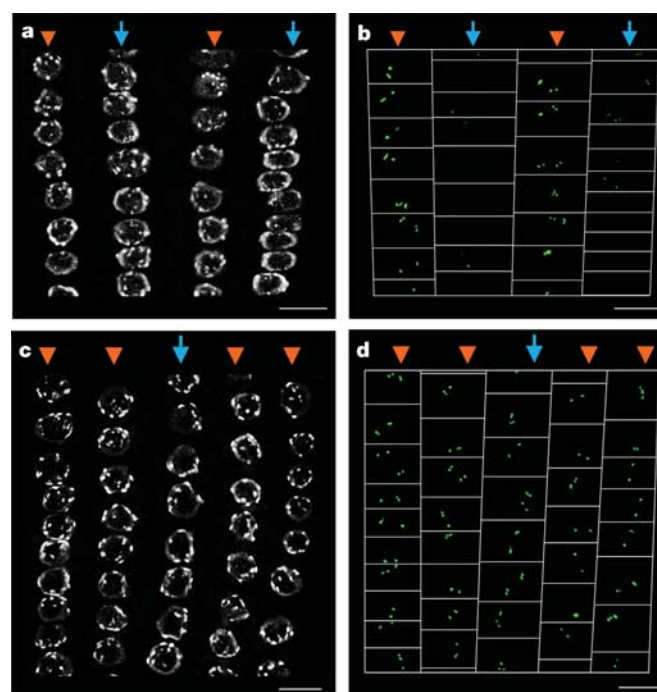


Figure 1 | Cell-type-specific chromatin organization at the *GL2* locus in meristematic epidermal nuclei. In all figures, orange arrowheads indicate atrichoblast nuclei; blue arrows indicate trichoblast nuclei; images are oriented with the root-tip down; and grids represent the location of cell walls. **a, b**, In wild type, the chromatin state differs between the two types of cell: FISH signals are strong in atrichoblasts, but are absent or very weak in trichoblasts. **a**, DAPI stain. **b**, FISH with *GL2*BAC probe (green). **c, d**, In the *cpc* mutant, the chromatin state in the *GL2* region is the same in all nuclei. **c**, DAPI stain. **d**, FISH with *GL2*BAC probe (green). Scale bars, 10 μ m.

¹Department of Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK.

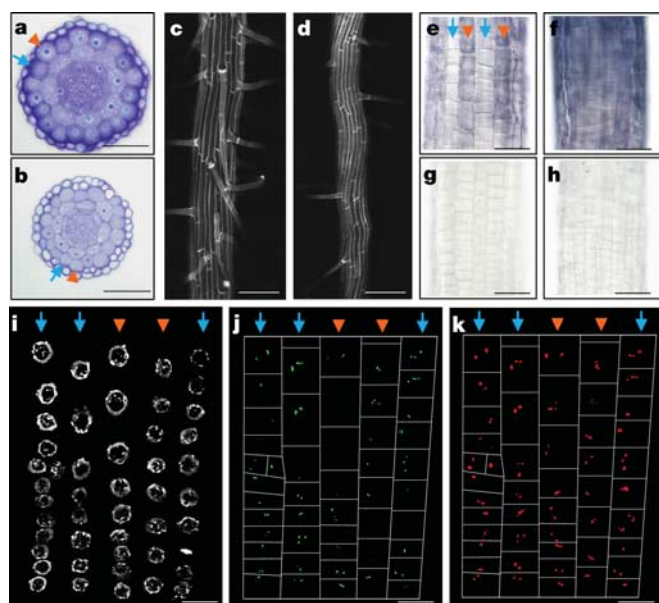


Figure 2 | FAS2 is required to control position-dependent cell-type specification. **a, b**, FAS2 acts early in epidermal cell fate specification. Shown are transverse sections in the meristematic region of wild-type (**a**) and *fas2* (**b**) seedling roots. The different cytoplasmic density and degree of vacuolation present in wild type between cells in the atrichoblast and those in the trichoblast position are lost in *fas2*. Scale bars, 50 μ m. **c, d**, Projections of confocal optical sections of the differentiated epidermis of wild-type (**c**) and *fas2* (**d**) seedling roots show the reduced number of hairs present in *fas2*. Scale bars, 100 μ m. On average, the wild type has 41 ± 6 hairs per mm and *fas2* has 20 ± 7 hairs per mm. **e–h**, *In situ* localization of *GL2* messenger RNA in the meristematic epidermis. **e**, In wild type, *GL2* is expressed in atrichoblast files. **f**, In *fas2*, the pattern expression of *GL2* is lost ($n = 100$). Wild type (**g**) and *fas2* (**h**) hybridized with a *GL2* sense probe show no signal. Scale bars, 30 μ m. **i–k**, FISH on *fas2* meristematic epidermal nuclei showing lack of patterned chromatin organization at the *GL2* locus. **i**, DAPI stain. **j**, FISH with *GL2*BAC probe (green). **k**, FISH with BAC probes (red) flanking the *GL2* region. Signals are present in every nucleus. Scale bars, 10 μ m.

every cell, regardless of its position relative to the underlying cortical cells (Fig. 1c, d, and Supplementary Figs 2c, d and 3a–d). These results show that CPC is required to establish the closed conformation in the trichoblasts of the epidermis, but that neither *GL2* expression nor cell fate specification is required to set up or to maintain an open chromatin state at the *GL2* locus.

We further investigated whether chromatin organization controls root epidermal cell-type specification by analysing the *fasciata2* mutant (*fas2*). In *fas2*, the activity of one of the three subunits of the chromatin assembly factor-1 (CAF-1) is defective, resulting in incorrect assembly of chromatin during replication¹⁴. We found that FAS2 is required for the correct establishment of root epidermal pattern (Fig. 2a–d). Hybridization experiments with the *GL2*BAC probe on *fas2* meristematic epidermal tissue showed that in *fas2* the patterned chromatin organization at the *GL2* locus observed in wild type is lost. In *fas2*, most nuclei are in an open chromatin state, but are located both in atrichoblast and trichoblast positions; the remaining nuclei are in a closed chromatin state and are also located in both positions (Fig. 2i, j). To confirm that patterns of chromatin organization are specific to the genomic region around *GL2*, we simultaneously labelled cells with two other BAC probes flanking each side of the *GL2* region. We observed clear FISH signals for these probes in every nucleus (Fig. 2i, k). This indicates that in *fas2* chromatin organization is affected only in the genomic region around the *GL2* locus, not in the flanking regions. Moreover, *GL2* is ectopically expressed in the *fas2* mutant background (Fig. 2e–h). In *fas2*, therefore, epidermal cells cannot specifically organize chromatin at the *GL2* locus in alternative states according to the

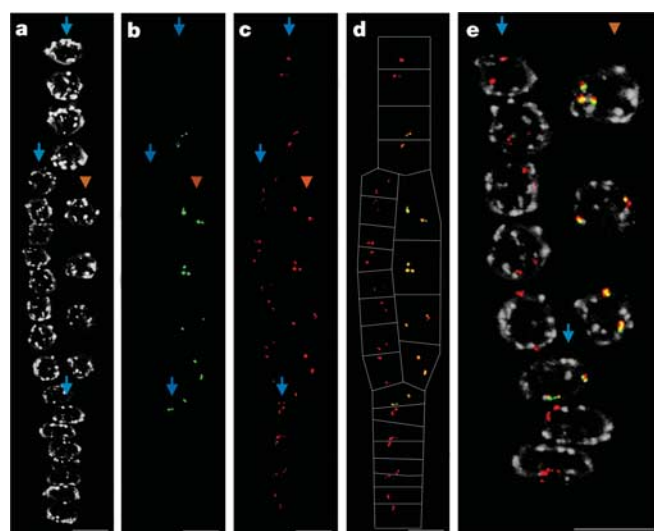


Figure 3 | Epidermal T clones show that chromatin state in the *GL2* region is defined by position and not by cell lineage. **a**, DAPI stain. **b**, FISH with *GL2*BAC probe (green). Signals are present only in atrichoblasts. **c**, FISH with BAC probes (red) flanking the *GL2* region. Signals are present in both atrichoblasts and trichoblasts. **d**, Superimposition of **b** and **c**. **e**, Close-up of merged images in **a**, **b** and **c**. Scale bars, 10 μ m.

underlying positional information, resulting in an inability to establish an epidermal pattern. This indicates that the correct assembly of chromatin is required to propagate the positional cues to chromatin state and to cell fate.

To analyse further the role of positional information in controlling chromatin organization and cell specification, we determined whether the patterns of chromatin organization in the *GL2* region can be remodelled in response to changes in positional information by examining root epidermal T clones. In T clones, parallel files of cells with the same clonal origin are exposed to different positional cues, differentially express *GL2* and acquire distinct fates⁶. In the T clones, bright signals were detected in the atrichoblast file with the *GL2*BAC probe, but signals were not detected in the trichoblast file (Fig. 3a, b). FISH signals were absent in trichoblasts belonging to the file from which the epidermal T clones originated, with the exception of those adjoining the T clones (Fig. 3a, b). This shows that the differential chromatin organization is not irreversibly fixed, but is remodelled in response to the new positional cues to which the cells in the T clones are exposed. We also simultaneously hybridized the T clone with the probes flanking the *GL2* region and observed clear FISH signals for these probes in every nucleus of both files in the T clones (Fig. 3a, c) as well as in the file from which the epidermal clones had originated (Fig. 3a, c). This indicates that the patterns of chromatin organization are specific for the genomic region around the *GL2* locus, and that only in this region is the chromatin remodelled on changes in positional information (Fig. 3d, e).

Having shown that chromatin organization is instrumental in controlling gene expression in a position-dependent manner, we determined how rapidly the chromatin state around the *GL2* locus is remodelled when changes in positional information occur. We examined a two-cell epidermal T clone—that is, a newly formed clone composed of only two cells derived from the longitudinal division of an epidermal cell and in which *GL2* is already differentially expressed between the cells⁶. In such a clone the *GL2*BAC signal was detected only in the atrichoblast, whereas the flanking signals were detected in both cells (Fig. 4a, b), indicating that the chromatin state at the *GL2* locus is remodelled immediately after the longitudinal division has occurred and progression through a whole cell cycle is not required.

To establish at what stage of the cell cycle changes in chromatin

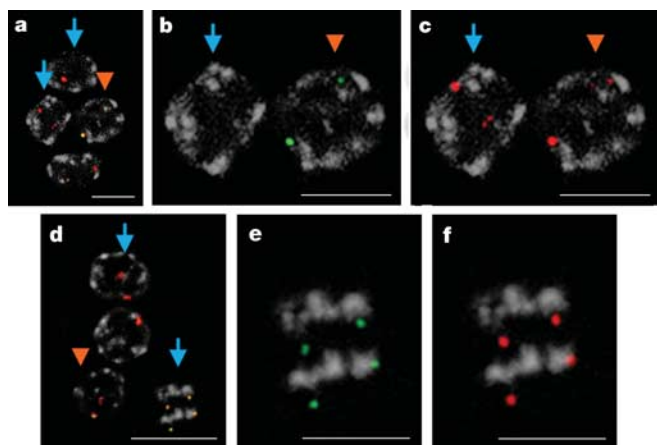


Figure 4 | Chromatin state in the *GL2* region is reset upon change in cell position. **a**, Two-cell epidermal T clone, counterstained with DAPI (grey) and probed with *GL2BAC* probe (green) and BAC probes (red) flanking the *GL2* region. **b**, Close-up of the two-cell epidermal T clone in **a**. FISH signals for the *GL2BAC* probe (green) are strong in the atrichoblast nucleus, but absent in the trichoblast nucleus. **c**, Two FISH signals plus one very strong signal, produced by two merged signals, are seen in both atrichoblast and trichoblast with the flanking probes (red), indicating that the cells are in the G1 phase of the cell cycle. **d–f**, An epidermal T clone showing a mitotic trichoblast in anaphase. Strong labelling with both the *GL2BAC* probe (green) and the flanking probes (red) shows that the chromatin in the *GL2* region is accessible at mitosis. Scale bars, 5 μ m.

organization occur, we counted the FISH signals detected in each nucleus of the two-cell epidermal T clone. *Arabidopsis* is a diploid species and epidermal meristematic nuclei show two FISH signals with a single copy probe when they are in G1 and four signals when in G2 (Supplementary Fig. 4). In the two-cell T clone, we observed numbers of FISH signals indicating that both nuclei are in G1 (Fig. 4a–c). Furthermore, the *GL2BAC* signal was clearly detected in mitotic figures in trichoblast cell files, in contrast to its absence in interphase (Fig. 4d–f and Supplementary Fig. 5), suggesting that the factors responsible for the closed conformation of this chromatin region during interphase may be removed or modified during mitosis. Thus, the chromatin state is reset at mitosis and is respecified during the following G1 phase according to the underlying positional information. This permits cell fate changes to occur rapidly with no need for DNA synthesis or completion of a whole cell cycle, indicating that cells are constantly assessing their ‘surroundings’ and responding to them by remodelling chromatin.

In animal and plant development, the correct interpretation of positional signals is important for proper tissue and organ formation^{15–19}. Our findings indicate that chromatin organization integrates positional cues with gene expression programmes to direct cell fate specification. We have shown that the chromatin state can be reset and remodelled each cell cycle during development. This capacity may be the basis of the plant cell’s ability to respond promptly to new stimuli and to change fate accordingly and may be important in ultimately understanding the great developmental plasticity of cells.

METHODS

Plant material and growth conditions. Wild-type Columbia-0 (Col-0) ecotype and *fas2* (*Ler*) seeds were obtained from the Nottingham Stock Centre, *cpc* mutant seeds (WS) were a gift from K. Okada (Kyoto University, Japan), and *wer-1* mutant seeds (Col) were a gift from J. Schiefelbein (University of Michigan, USA). Seedlings were grown for 3 d as described¹.

Microscopy. Analysis of the radial organization of the root epidermis in the meristematic zone was done as described¹ on transverse sections of roots of 3-day-old wild-type ($n = 6$) and *fas2* ($n = 11$) seedlings embedded in Technovit resin (Heraeus). Root hair numbers were counted on 10- μ m serial transverse sections taken from 3-day-old wild-type ($n = 6$) and *fas2* ($n = 11$) seedlings

embedded in Technovit resin. Confocal images of propidium-iodide-stained 3-day-old wild-type and *fas2* seedlings were collected as described⁸.

DNA *in situ* hybridization. Seedlings were fixed overnight in 4% (w/v) formaldehyde plus 0.1% Tween-20 at 4 °C, and rinsed twice in 1 $\times$ PBS plus 0.1% Tween-20 and twice in distilled H₂O. Roots were then manually dissected under a stereomicroscope and aligned onto poly-lysine-treated slides and left to dry overnight at 42 °C. The material was permeabilized for 40 min at 30 °C in 1% (w/v) driselase, 2% (w/v) cellulase and 0.3% (w/v) pectolyase in 1 $\times$ PBS, washed twice in 1 $\times$ PBS and once in 2 $\times$ SSC, treated with RNase (10 μ g ml⁻¹ in 2 $\times$ SSC) for 1 h at 37 °C, and washed twice in 2 $\times$ SSC. Slides were equilibrated in pronase buffer (100 mM Tris-HCl and 10 mM EDTA; pH 7.5), treated with pronase (100 μ g ml⁻¹) for 15 min at room temperature, stopped with 0.2% glycine in 1 $\times$ PBS for 2 min, washed in 1 $\times$ PBS, treated with 0.5% (v/v) acetic anhydride in 0.1 M triethanolamine for 10 min, and then washed twice in 1 $\times$ PBS. Probes were labelled with biotin-dUTP or digoxigenin-dUTP (Boehringer Mannheim) by nick translation. BAC clone F19K16, which contains an insert of 111 kb, was used as a probe for the *GL2* region (*GL2BAC* probe). BAC clones F18B13, with a 117-kb insert, and F20B17, with a 90-kb insert, were used as probes for the genomic regions flanking the *GL2* region. BAC F18B13 overlaps the 5' end of BAC F19K16 by 41 kb, and BAC F20B17 overlaps the 3' end of BAC F19K16 by 35 kb. The hybridization mixture (7 ng μ l⁻¹ of labelled DNA, 50% formamide, 10% dextran sulphate, 2 $\times$ SSC and 0.125% SDS) was applied to the slides, which were then denatured at 75 °C for 2 min and hybridized overnight at 37 °C. After hybridization, slides were washed at 42 °C once in 2 $\times$ SSC, twice in 20% formamide plus 0.1 $\times$ SSC and twice in 2 $\times$ SSC, and then twice in 2 $\times$ SSC at room temperature and twice in 4 $\times$ SSC plus 0.2% Tween-20. Biotin-dUTP probes were detected with Cy3-conjugated antibodies (Sigma), digoxigenin-dUTP probes were detected with fluorescein isothiocyanate (FITC)-conjugated antibodies (Sigma), nuclei were counterstained with DAPI (4',6-diamidino-2-phenylindole dihydrochloride hydrate; Sigma; 1 μ g ml⁻¹), and slides were mounted in Vectashield (Vector Laboratories).

RNA *in situ* hybridization was done on 3-day-old wild-type and *fas2* seedling roots essentially as described⁸.

Digital microscopy, data collection, processing and analysis. Samples were imaged with a Nikon Eclipse 600 microscope equipped with a Hamamatsu ORCA-ER cooled CCD camera and a Prior Proscan x, y, z stage. z-series sections at a separation of 0.25–0.5 μ m were collected with Metamorph software (Universal Imaging Corporation), deconvolved with Autodeblur (Autoquant Inc.) and processed with ImageJ (<http://rsb.info.nih.gov/ij/>) and Photoshop7.

The signal intensity in FISH experiments was measured with ImageJ by calculating the integrated intensity in a 64 $\times$ 64 pixel square box around the signal in individual nuclei in trichoblast and atrichoblast positions. Signal intensities were compared only between trichoblast and atrichoblast nuclei within the same images of intact tissue samples, so that reliable quantitative comparisons could be made. In the scatter plots, the signal intensities of epidermal root nuclei were represented by plotting the values of each atrichoblast and trichoblast relative to the mean intensity of all of the epidermal nuclei.

Images shown in the figures are maximum intensity projections of series of optical z sections. The identities of the files of nuclei (that is, atrichoblast or trichoblast) were defined on the basis of their position relative to the underlying cortical nuclei.

In trichoblasts, no signal is detected with the *GL2BAC* probe, meaning that its target sequence is inaccessible for its whole length, whereas signal is detected with the two flanking probes, which each overlap the *GL2BAC* for about a third of their length. The fact that we see a signal for the flanking BACs shows that the target sequences of the non-overlapping, greater part of each of these BACs are accessible.

Received 6 June; accepted 27 September 2005.

Published online 18 December 2005.

1. Dolan, L. *et al.* Clonal relationships and patterning in the root epidermis of *Arabidopsis*. *Development* **120**, 2465–2474 (1994).
2. Galway, M. E. *et al.* The *TTG* gene is required to specify epidermal cell fate and cell patterning in the *Arabidopsis* root. *Dev. Biol.* **166**, 740–754 (1994).
3. Masucci, J. D. *et al.* The homeobox gene *GLABRA2* is required for position-dependent cell differentiation in the root epidermis of *Arabidopsis thaliana*. *Development* **122**, 1253–1260 (1996).
4. Di Cristina, M. *et al.* The *Arabidopsis* *Athb-10* (*GLABRA2*) is an HD-Zip protein required for regulation of root hair development. *Plant J.* **10**, 393–402 (1996).
5. Hung, C. Y. *et al.* A common position-dependent mechanism controls cell-type patterning and *GLABRA2* regulation in the root and hypocotyl epidermis of *Arabidopsis*. *Plant Physiol.* **117**, 73–84 (1998).

6. Berger, F., Haseloff, J., Schiefelbein, J. & Dolan, L. Positional information in the root epidermis is defined during embryogenesis and acts in domains with strict boundaries. *Curr. Biol.* **8**, 421–430 (1998).
7. Lin, Y. & Schiefelbein, J. Embryonic control of epidermal cell patterning in the root and hypocotyl of *Arabidopsis*. *Development* **128**, 3697–3705 (2001).
8. Costa, S. & Dolan, L. Epidermal patterning genes are active during embryogenesis in *Arabidopsis*. *Development* **130**, 2839–2901 (2003).
9. Kwak, S. H., Shen, R. & Schiefelbein, J. Positional signalling mediated by a receptor-like kinase in *Arabidopsis*. *Science* **307**, 1111–1113 (2005).
10. Lee, M. M. & Schiefelbein, J. WEREWOLF, a MYB-related protein in *Arabidopsis*, is a positional-dependent regulator of epidermal cell patterning. *Cell* **99**, 473–483 (1999).
11. Lee, M. M. & Schiefelbein, J. Cell pattern in the *Arabidopsis* root epidermis determined by lateral inhibition with feedback. *Plant Cell* **14**, 611–618 (2002).
12. Wada, T., Tachibana, T., Shimura, Y. & Okada, K. Epidermal cell differentiation in *Arabidopsis* determined by a *Myb* homolog, *CPC*. *Science* **277**, 1113–1116 (1997).
13. Wada, T. *et al.* Role of a positive regulator of root hair development, *CAPRICE*, in *Arabidopsis* root epidermal cell differentiation. *Development* **129**, 5409–5419 (2002).
14. Kaya, H. *et al.* FASCIATA genes for chromatin assembly factor-1 in *Arabidopsis* maintain the cellular organization of apical meristems. *Cell* **104**, 131–142 (2001).
15. Wolpert, L. Cell boundaries: knowing who to mix with and what to shout or whisper. *Development* **130**, 4497–4500 (2003).
16. Szymkowiak, E. J. & Sussex, I. M. What chimeras can tell us about plant development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 351–376 (1996).
17. Dolan, L. & Okada, K. Signalling in cell specification. *Semin. Cell Dev. Biol.* **10**, 149–156 (1999).
18. Kidner, C., Sundaresan, V., Roberts, K. & Dolan, L. Clonal analysis of the *Arabidopsis* root confirms that position, not lineage, determines cell fate. *Planta* **211**, 191–199 (2000).
19. Laux, T. The stem cell concept in plants: a matter of debate. *Cell* **113**, 281–283 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Dolan, G. Moore, C. Dean and M. Towers for discussion and advice. This work was funded by the BBSRC and the Gatsby Charitable Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.S. (peter.shaw@bbsrc.ac.uk).

A phosphatase complex that dephosphorylates γ H2AX regulates DNA damage checkpoint recovery

Michael-Christopher Keogh^{1*}, Jung-Ae Kim^{4*}, Michael Downey^{5,7*}, Jeffrey Fillingham^{5,6}, Dipanjan Chowdhury^{2,8}, Jacob C. Harrison⁴, Megumi Onishi³, Nira Datta^{5,6}, Sarah Galicia⁷, Andrew Emili^{5,6}, Judy Lieberman^{2,8}, Xuetong Shen⁹, Stephen Buratowski¹, James E. Haber⁴, Daniel Durocher^{5,7}, Jack F. Greenblatt^{5,6} & Nevan J. Krogan^{5,6†}

One of the earliest marks of a double-strand break (DSB) in eukaryotes is serine phosphorylation of the histone variant H2AX at the carboxy-terminal SQE motif to create γ H2AX-containing nucleosomes¹. Budding-yeast histone H2A is phosphorylated in a similar manner by the checkpoint kinases Tel1 and Mec1 (ref. 2; orthologous to mammalian ATM and ATR, respectively) over a 50-kilobase region surrounding the DSB³. This modification is important for recruiting numerous DSB-recognition and repair factors to the break site, including DNA damage checkpoint proteins^{4,5}, chromatin remodellers⁶ and cohesins^{7,8}. Multiple mechanisms for eliminating γ H2AX as DNA repair completes are possible, including removal by histone exchange followed potentially by degradation, or, alternatively, dephosphorylation. Here we describe a three-protein complex (HTP-C, for histone H2A phosphatase complex) containing the phosphatase Pph3 that regulates the phosphorylation status of γ H2AX *in vivo* and efficiently dephosphorylates γ H2AX *in vitro*. γ H2AX is lost from chromatin surrounding a DSB independently of the HTP-C, indicating that the phosphatase targets γ H2AX after its displacement from DNA. The dephosphorylation of γ H2AX by the HTP-C is necessary for efficient recovery from the DNA damage checkpoint.

If γ H2AX is dephosphorylated by a phosphoserine-specific phosphatase, it should accumulate when the relevant phosphatase is eliminated. There are about 67 phosphatases in *Saccharomyces cerevisiae*. Of the 62 not essential for cell viability, 22 are predicted to be phosphoserine protein phosphatases, and, of these, nine seem to be nuclear⁹. We tested the corresponding deletion strains by western analysis with the use of antibody specific for the phosphorylated form of histone H2A (Fig. 1a and Supplementary Fig. S1a). Only with deletion of *PPH3/YDR075w* did γ H2AX increase significantly. This phosphorylation event occurs primarily in S phase (Supplementary Fig. S2a) and is dependent on *MEC1* and *TEL1* activity (Supplementary Fig. S2b), thus probably reflecting γ H2AX accumulation around DSBs that occur spontaneously during DNA replication.

To characterize this putative γ H2AX phosphatase further, we used a strain containing a C-terminal tandem affinity purification (TAP) tag¹⁰ on Pph3. After affinity purification we identified components of the highly purified protein complex by digestion with trypsin, followed by mass spectrometry¹¹. Pph3 purified together with roughly stoichiometric amounts of two polypeptides, Ynl201c/Psy2 and Ybl046w (Fig. 1b). Tagging and purification of these two

proteins, in turn, identified the same complex: the HTP-C. Deletion of *PSY2* or *YBL046W* also increased the cellular levels of γ H2AX with or without DNA damage by methyl methane sulphonate (MMS) (Fig. 1c), indicating that all three proteins might work as one functional unit *in vivo*. By 4 h after removal of the MMS, γ H2AX levels had returned to background in wild-type cells but remained high in *pph3Δ* cells (data not shown).

To show that the HTP-C could regulate γ H2AX levels directly, we phosphorylated H2A *in vitro* (Supplementary Fig. S1b) and performed dephosphorylation assays with IgG-precipitated Pph3-TAP. As controls, we used immunoprecipitates of two other related and nuclear-localized, TAP-tagged phosphatases, Pph21 and Pph22 (Supplementary Fig. S1c), deletions of which have no effect on γ H2AX levels (Fig. 1a). Small amounts of Pph3 efficiently dephosphorylated γ H2AX *in vitro* (Fig. 1d), especially in the presence of Mn^{2+} rather than Mg^{2+} (Supplementary Fig. S1d), whereas much larger amounts of Pph21 or Pph22 had only marginal activity on this substrate. Therefore, *in vivo* and *in vitro* data both indicate that the HTP-C dephosphorylates γ H2AX. The metazoan homologue of the HTP-C is currently unclear: the PPP4c phosphatase exists in an analogous complex¹², whereas the PP2A phosphatase regulates γ H2AX levels *in vivo*¹³.

To obtain further evidence that the HTP-C is related to DNA repair, we used synthetic genetic array technology¹⁴ to create double mutants that were scored for growth and DNA damage sensitivity. Tetrad dissection showed that deletion of *PPH3* caused synthetic (synergistic) growth defects in combination with deletions of genes coding for homologous recombination (HR) proteins (*RAD50*, *MRE11*, *RAD51*, *RAD52* and *SAE2*), the Flap endonuclease Rad27, the chromatin assembly factor Asf1, the DNA helicase Rrm3 and replication-fork-associated components (*TOF1*, *CSM3*, *MRC1*, *CTF8*, *CTF18* and *DCC1*) (Fig. 2a, and Supplementary Fig. S3). The fact that *pph3Δ* displayed slow-growth phenotypes when combined with deletions of DNA replication and repair genes argues that HTP-C activity is important for one or more aspects of the cellular DNA damage response. HTP-C deletions alone did not cause sensitivity to the DNA-damaging agents camptothecin, bleomycin, hydroxyurea or MMS, but double deletions such as *pph3Δ ctf18Δ* exhibited hypersensitivity to these drugs (Fig. 2b, Supplementary Fig. S3, and data not shown). We note that deletions of *PSY2* and *PPH3* were also found recently to confer sensitivity to the DNA-damaging agent cisplatin¹⁵.

¹Department of Biological Chemistry and Molecular Pharmacology, ²Department of Pediatrics, and ³Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Rosenstiel Center and Department of Biology, Brandeis University, Waltham, Massachusetts 02454, USA. ⁵Departments of Medical Genetics and Microbiology, and ⁶The Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L6, Canada. ⁷Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. ⁸CBR Institute for Biomedical Research, Boston, Massachusetts 02115, USA. ⁹Department of Carcinogenesis, MD Anderson Cancer Center, Smithville, Texas 78957, USA. [†]Present address: Department of Cellular and Molecular Pharmacology, UCSF, San Francisco, California 94143, USA.

*These authors contributed equally to this work.

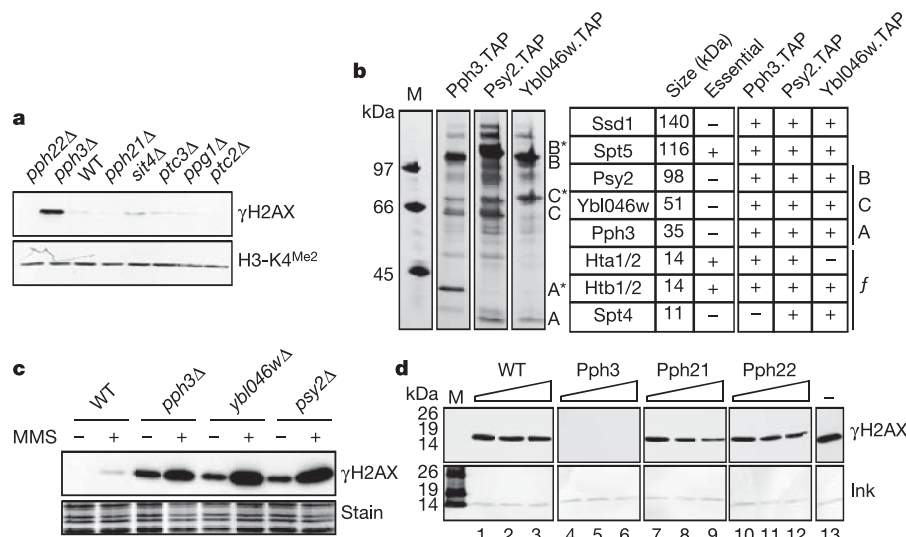


Figure 1 | The HTP-C regulates H2A *in vivo* and dephosphorylates γH2AX *in vitro*. **a**, γH2AX antibody immunoreactivity of whole cell extracts from phosphatase deletion cells (see also Supplementary Fig. S1). Lower panel, loading control. WT, wild type. **b**, Identification of Pph3-associated factors. Pph3 (A for the native protein; A* for the calmodulin binding peptide (CBP)-tagged form) co-purified with Psy2 (B, B*), Ybl046w (C, C*), and substoichiometric amounts of a putative RNA exonuclease (Ssd1), Spt4/Spt5 and histones H2A (Hta1/2) and H2B (Htb1/2) (see also Supplementary Fig. S4). *f* indicates smaller proteins run off the gel and

detected by liquid chromatography/mass spectrometry. M, molecular mass marker proteins. **c**, Deletions of *PPH3*, *PSY2* or *YBL046W* increase γH2AX levels in undamaged or damaged (0.05% MMS) cells. Lower panel, loading control for total protein. **d**, HTP-C is an efficient γH2AX phosphatase *in vitro*. Reactions contained γH2AX/H2B dimer substrate and immunoprecipitates from an untagged wild type or the indicated epitope-tagged phosphatase strains. γH2AX was detected by western blotting. Lower panel, loading control for total H2A/H2B.

S. cerevisiae uses the homothallic (HO) endonuclease to switch mating types by creating a DSB in *MAT* and repairing the break by HR with donor sequences at *HML* or *HMR*¹⁶. Expressing HO from a galactose-inducible promoter and inserting its 24-base-pair recognition sequence at different locations provides an opportunity to monitor DSB recognition and repair¹⁷. In the experiments shown in Fig. 3a, b, HR occurred *in trans* through an uncuttable *MATa*-inc donor on chromosome V in strains lacking *HML* and *HMR* (Supplementary Table S1)^{18,19}. HO cleavage was essentially complete within 0.5 h, repair products were detectable within 2 h, and more than 80% of cells successfully completed HR (Fig. 3a, and data not shown). After HO cleavage, γH2AX appeared around the DSB and covered about 50 kilobases (kb) in less than an hour (data not shown). After about 2 h, H2A phosphorylation around the DSB began to decline (as detected by chromatin immunoprecipitation (ChIP)), with the loss at 5 kb from the DSB slightly earlier than at 10 kb, and significantly earlier than at 20 kb (Fig. 3b, and data not shown). Levels of γH2AX 5 and 10 kb from the DSB declined before the final repair product appeared (detected by Southern blotting; Fig. 3a, b). Indeed, the kinetics of loss was correlated with the appearance of an early gene conversion intermediate, the beginning of new DNA synthesis from the invading 3' end of the recipient strand on the donor strand template (strand invasion–primer extension (SI–PE); Fig. 3b)²⁰. This indicates that the signal to trigger γH2AX loss might not be the completion of DSB repair but rather the completion of a step that will normally lead to repair. Note that a HR-defective *rad54Δ* strain, which fails at this SI–PE step²¹, has a normal γH2AX–ChIP profile (Supplementary Fig. S5), again indicating that the loss of γH2AX occurs after synapse formation but before the completion of repair.

DSB repair in *pph3Δ* cells was as efficient as in wild-type cells and followed the same kinetics (Fig. 3a, b). Moreover, the loss of γH2AX from the DSB region in ChIP experiments occurred with rates similar to those in wild-type cells (Fig. 3a, b). Note that a high ChIP background reduced the fold increase in relative γH2AX levels at the DSB in *pph3Δ* cells, possibly because the phosphorylated histone was recycled into chromatin (see below, and Supplementary Figs S6

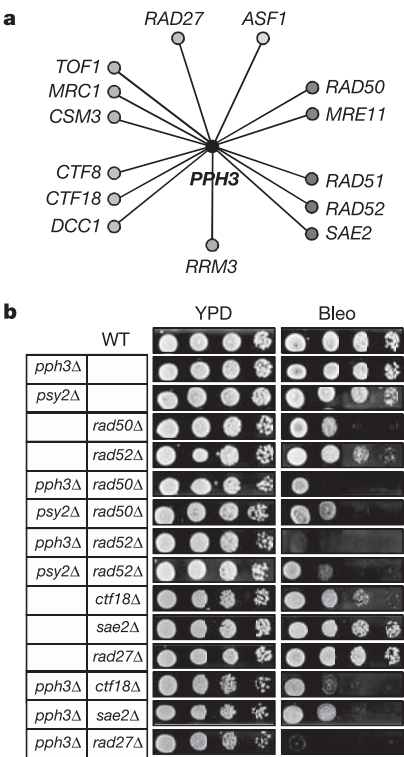


Figure 2 | The HTP-C interacts genetically with DNA repair factors. **a**, Lines connect genes with synthetic genetic interactions confirmed by tetrad dissection (Supplementary Fig. S3, and data not shown). **b**, Fivefold serial dilutions of various double-mutant strains on YPD plates with or without bleomycin (Bleo). WT, wild type.

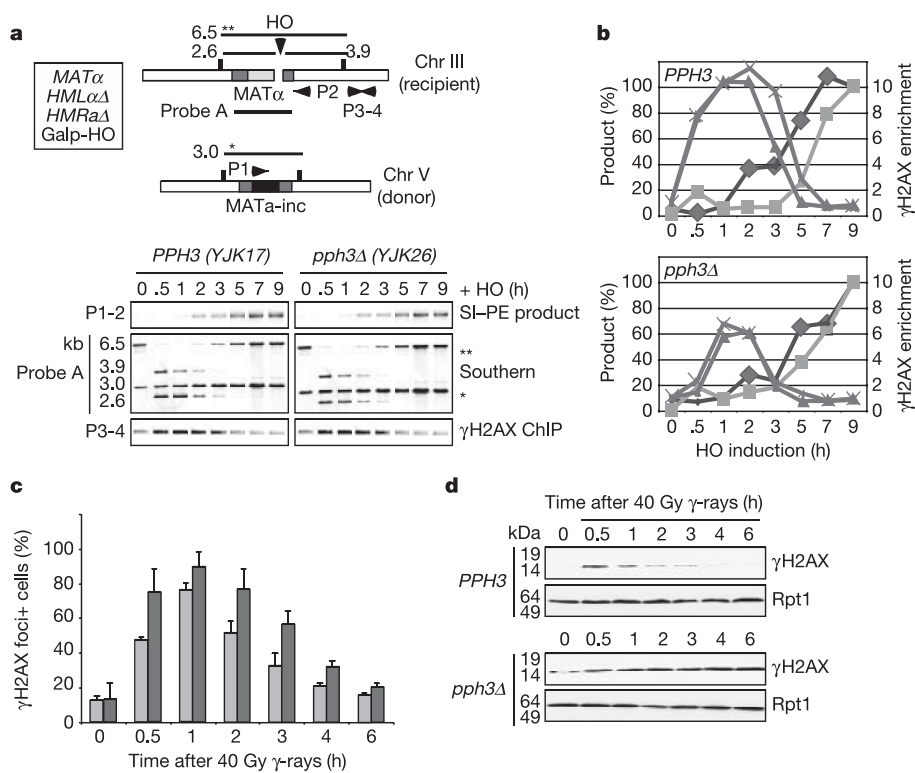


Figure 3 | The HTP-C is not required for DSB repair and dephosphorylates displaced γ H2AX rather than chromatin-associated γ H2AX at a DSB. **a**, Pph3 is not required for γ H2AX removal or efficient DSB repair by ectopic recombination. The top diagram shows the ectopic recombination system. Chr, chromosome. SI-PE (primers P1 and P2) quantifies synapse formation with donor DNA sequence. Southern blotting with probe A after digestion with *EcoRI* monitors recipient MAT α (uncut 6.5 kb (two asterisks), HO-cut 2.6 and 3.9 kb), donor MATa-inc (3.0 kb (asterisk), loading control) and major repair product (6.5 kb (two asterisks)). ChIP with anti- γ H2AX 5 kb from the break site (primers P3 and P4). **b**, γ H2AX at a DSB disappears before completion of repair and independently of Pph3. SI-PE (diamonds) and Southern (squares) analyses in **a** are quantified relative to final repair products 9 h after HO induction. ChIP analyses are expressed as γ H2AX levels 5 kb (triangles) or 10 kb (crosses) from the DSB relative to an undamaged DNA region. **c**, γ H2AX foci disassembly after γ -irradiation does not require the HTP-C (200 cells counted per time point; mean and s.e.m. calculated from two independent time courses). Light grey bars, PPH3; dark grey bars, *pph3* Δ . **d**, Western blotting of γ H2AX after γ -irradiation in wild-type or *pph3* Δ cells. Lower panel, loading control.

and S9). We confirmed this result by staining with anti- γ H2AX antibody to follow the kinetics of DSB repair foci after γ -irradiation (Fig. 3c, d). γ H2AX foci were efficiently disassembled by 6 h after damage in *pph3* Δ cells (Fig. 3c) even though H2A remained

phosphorylated (Fig. 3d). These results indicate that γ H2AX dephosphorylation might occur mainly after γ H2AX/H2B dimers are displaced (Supplementary Fig. S4), perhaps by a chromatin remodeller analogous to the *Drosophila* Tip60 complex²². γ H2AX

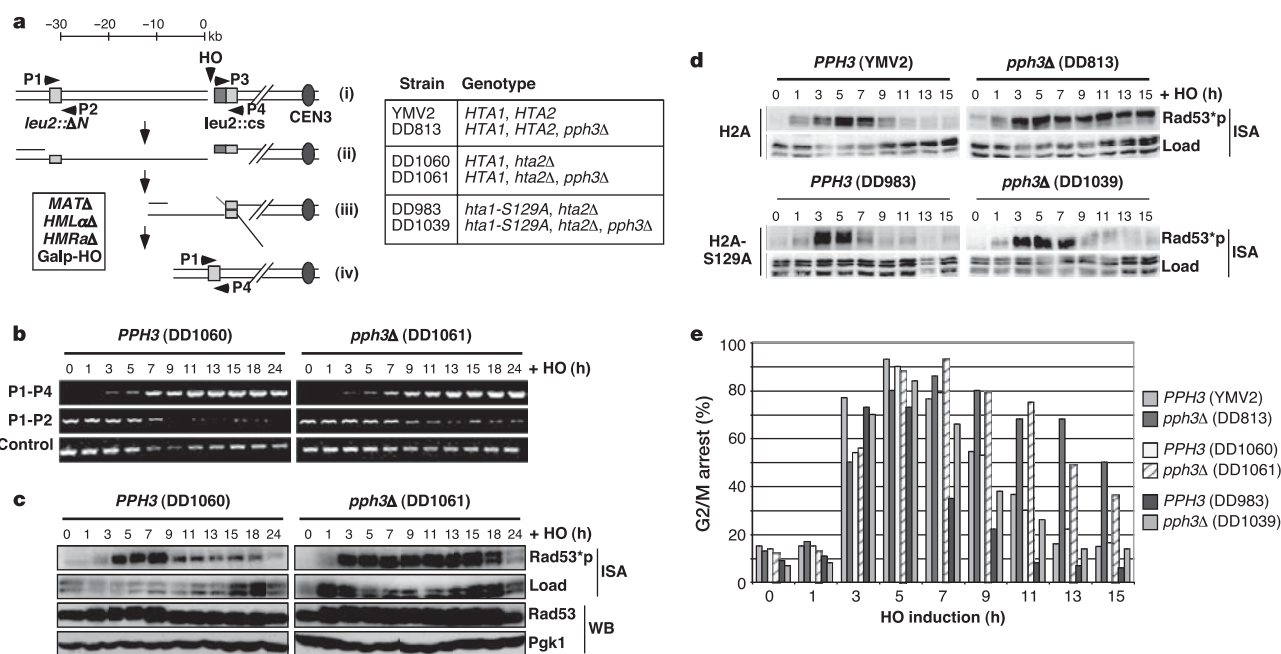


Figure 4 | γ H2AX dephosphorylation by the HTP-C is required for DNA damage checkpoint recovery. **a**, Diagram for repair by SSA of an HO-induced DSB. Top: galactose-induced HO cuts *leu2::cs*. Second row: processive 5' $\rightarrow$ 3' resection (4 kb h⁻¹) generates ssDNA and eventually reaches *leu2::N*. Third row: the two homologous regions anneal in a Rad52- and Rad1/10-dependent fashion and the non-homologous sequences are excised. Bottom: after successful SSA the damage checkpoint is relieved¹⁹. P1–P4 are primers used for PCR analyses. Reduced histone H2A dosage has no effect on these analyses (compare YMV2 and DD1060 strains). **b**, SSA

occurs between 7 and 9 h in both wild-type and *pph3* Δ strains. This agrees with the estimated repair timing based on resection rate (8–9 h). **c**, Persistent checkpoint activation in *pph3* Δ cells monitored by Rad53 *in situ* autophosphorylation (ISA) (top panels) or mobility in western blotting (WB)³⁰. Autophosphorylation of a non-DNA damage inducible kinase is used as a loading control. **d**, Rad53 activation does not persist in *pph3* Δ *hta1-S129A* cells. **e**, Pph3 is required for removal of the cell cycle checkpoint arrest maintained by γ H2AX. The graph shows the percentage of cells in G2/M arrest various times after HO induction.

is likely to be actively removed: at both 10 and 20 kb from the DSB, most of the DNA remained double-stranded after DSB repair (Supplementary Fig. S5), indicating that γ H2AX loss is not mediated by single-stranded DNA (ssDNA) resection.

To assess whether the DNA damage checkpoint was affected by deleting *PPH3*, we examined checkpoint activation and termination kinetics in isogenic *PPH3* and *pph3 Δ* strains with the use of the YMV2 system (Fig. 4a, and Supplementary Table S1)¹⁹. In this background *MAT*, *HML* and *HMR* are deleted, and a galactose-inducible HO cuts at an engineered site in the *LEU2* gene. The induced DSB is repaired by single-stranded annealing (SSA)²³ from a partial duplication of *LEU2* located 30 kb away. This repair takes about 6–7 h, during which the ssDNA formed by 5' to 3' resection triggers DNA damage checkpoint-mediated G2/M arrest and phosphorylation of the yeast checkpoint adaptor Rad9 (refs 24, 25) and Chk2 homologue checkpoint kinase, Rad53 (refs 26, 27).

PPH3 and *pph3 Δ* cells were equally efficient in repairing the DSB (Fig. 4b), and the early kinetics of Rad53 and Rad9 activation after DSB induction was similar in both populations (Fig. 4b, top panels, and Supplementary Fig. S8c), peaking 5 h after DSB induction. However, whereas Rad53 and Rad9 activation decreased in wild-type cells 7–9 h after creation of the DSB, they remained active at least 8–9 h longer in the absence of the HTP-C (Fig. 4c, and Supplementary Fig. S8c), although these latter cells eventually recovered by 24 h. The sustained Rad53 and Rad9 activity was correlated with maintenance of the G2/M arrest (Fig. 4e). We therefore conclude that checkpoint recovery is profoundly delayed in *pph3 Δ* strains. When DNA repair is prevented, cells can escape long checkpoint-mediated G2/M arrest by adaptation²⁸. Whereas the known recovery mutants *srs2 Δ* , *ptc2 Δ* and *ptc3 Δ* are also defective in checkpoint adaptation^{19,29}, *pph3 Δ* cells show no deficiency in the timing of the latter process (Supplementary Fig. S7).

To ascertain whether the checkpoint recovery defect of the *pph3 Δ* strain was a direct consequence of the increased γ H2AX levels, we mutated the H2A phosphorylation site, Ser129, to Ala to yield *hta1-S129A* and *pph3 Δ hta1-S129A* mutant strains (Fig. 4a). Mutation of H2A Ser129 fully restored the ability of *pph3 Δ* strains to turn off checkpoint signalling in a timely manner (Fig. 4d, bottom panels) and re-enter the cell cycle (Fig. 4e) after DNA repair. Indeed, *hta1-S129A* strains turn off checkpoint signalling (Fig. 4d) and escape G2/M arrest (Fig. 4e) slightly earlier than their wild-type counterparts. The checkpoint recovery defect imparted by the deletion of *PPH3* therefore results from the persistent phosphorylation of H2A, and it is necessary for γ H2AX dephosphorylation to link the DNA repair process with the termination of checkpoint signalling. One intriguing possibility is that extensive formation of γ H2AX for more than 50 kb around a DSB keeps the DNA damage checkpoint active until the break has been fully repaired and the modified histone has been removed from the break site and dephosphorylated by the HTP-C.

METHODS

Full experimental details and protocols are available in the Supplementary Information. Descriptions of the yeast strains used to monitor DSB repair by ectopic recombination or single-strand annealing are found in the Supplementary Information and Supplementary Table 1. These strains employed inducible HO endonuclease to make a specific break. ChIP, Southern, SI-PE, single-strand resection, western or G2/M arrest analyses were performed in parallel as indicated.

Immunofluorescence microscopy and Western analyses were performed in parallel after DSB induction by γ -irradiation (40 Gy).

Double-mutant haploid strains containing both HTP-C and DNA metabolism deletions were created by synthetic genetic array technology. These double mutants were analysed for synthetic genetic interactions and sensitivity to DNA genotoxins (bleomycin, hydroxyurea or camptothecin).

In vitro phosphatase experiments used affinity-purified, TAP-tagged Pph3, Pph21 or Pph22 complexes. The immunopurified phosphatase was incubated with γ H2AX previously created *in vitro* by phosphorylating recombinant yeast

H2A/H2B dimers with RSK1. Phosphatase activity was measured by reactivity with anti- γ H2AX antibody.

Received 19 June; accepted 1 November 2005.

Published online 20 November 2005.

- Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. P. DNA double-strand breaks induce H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858–5868 (1998).
- Downs, J. A., Lowndes, N. F. & Jackson, S. P. A role for *Saccharomyces cerevisiae* histone H2A in DNA repair. *Nature* **408**, 1001–1004 (2000).
- Shroff, R. et al. Distribution and dynamics of chromatin modification at a defined DNA double strand break. *Curr. Biol.* **14**, 1703–1711 (2004).
- Nakamura, T. M., Du, L. L., Redon, C. & Russell, P. Histone H2A phosphorylation controls Crb2 recruitment at DNA breaks, maintains checkpoint arrest, and influences DNA repair in fission yeast. *Mol. Cell. Biol.* **24**, 6215–6230 (2004).
- Fernandez-Capetillo, O. et al. DNA-damage induced G2-M checkpoint activation by histone H2AX and 53BP1. *Nature Cell Biol.* **4**, 993–997 (2002).
- Thiriet, C. & Hayes, J. J. Chromatin in need of a fix: phosphorylation of H2AX connects chromatin to DNA repair. *Mol. Cell* **18**, 617–622 (2005).
- Ström, L., Lindroos, H. B., Shirahige, K. & Sjögren, C. Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. *Mol. Cell* **16**, 1003–1015 (2004).
- Ünal, E. et al. DNA damage response pathway uses histone modification to assess a double-strand break-specific cohesin domain. *Mol. Cell* **16**, 991–1002 (2004).
- Huh, W.-K. et al. Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
- Rigaut, G. et al. A generic protein purification method for protein complex characterisation and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
- Krogan, N. J. et al. RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell. Biol.* **22**, 6979–6992 (2002).
- Gingras, A. C. et al. A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol. Cell Proteomics* ahead of print publication, 2005 (doi:10.1074/mcp.M500231-MCP200).
- Chowdhury, D. et al. γ H2AX dephosphorylation by protein phosphatase 2A facilitates DNA double strand break repair. *Mol. Cell* (in the press).
- Tong, A. H. Y. et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
- Wu, H. I., Brown, J. A., Dorie, M. J., Lazzaroni, L. & Brown, J. M. Genome-wide identification of genes conferring resistance to the anticancer agents Cisplatin, Oxaliplatin and Mitomycin C. *Cancer Res.* **64**, 3940–3948 (2004).
- Haber, J. E. Mating-type gene switching in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* **32**, 561–599 (1998).
- Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).
- Moore, J. K. & Haber, J. E. Cell-cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 2164–2173 (1996).
- Vaze, M. B. et al. Recovery from checkpoint-mediated arrest after repair of a double-strand break requires Srs2 helicase. *Mol. Cell* **10**, 373–385 (2002).
- White, C. I. & Haber, J. E. Intermediates of recombination during mating type switching in *Saccharomyces cerevisiae*. *EMBO J.* **9**, 663–673 (1990).
- Sugawara, N., Wang, X. & Haber, J. E. *In vivo* roles of Rad52, Rad54 and Rad55 proteins in Rad51-mediated recombination. *Mol. Cell* **12**, 209–219 (2003).
- Kusch, T. et al. Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. *Science* **306**, 2084–2087 (2004).
- Fishman-Lobell, J., Rudin, N. & Haber, J. E. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol. Cell. Biol.* **12**, 1292–1303 (1992).
- Lee, S. E. et al. *Saccharomyces* Ku70, Mre11/Rad50 and RPA proteins regulate adaptation to G2/M arrest after DNA damage. *Cell* **94**, 399–409 (1998).
- Vialard, J. E., Gilbert, C. S., Green, C. M. & Lowndes, N. F. The budding yeast Rad9 checkpoint protein is subjected to Mec1/Tel1-dependent hyperphosphorylation and interacts with Rad53 after DNA damage. *EMBO J.* **17**, 5679–5688 (1998).
- Matsuoka, S., Huang, M. & Elledge, S. J. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science* **282**, 1893–1897 (1998).
- Blasina, A. et al. A human homologue of the checkpoint kinase Cds1 directly inhibits Cdc25 phosphatase. *Curr. Biol.* **9**, 1–10 (1999).
- Sandell, L. L. & Zakian, V. A. Loss of a yeast telomere: arrest, recovery and chromosome loss. *Cell* **75**, 729–739 (1993).
- Leroy, C. et al. PP2C phosphatases Ptc2 and Ptc3 are required for DNA checkpoint inactivation after a double-strand break. *Mol. Cell* **11**, 827–835 (2003).
- Pellucchi, A., Lee, S. E., Lucca, C., Foiani, M. & Haber, J. E. Regulation of *Saccharomyces* Rad53 checkpoint kinase during adaptation from DNA damage-induced G2/M arrest. *Mol. Cell* **7**, 293–300 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. J. Ingles for critically reading the manuscript and for discussion; C. Redon, W. Hofer and N. Lowndes for antibodies; E. O'Shea and J. Weissman for yeast strains; and X. Wu, G. Zhong and X. Guo for technical assistance. N.J.K. was supported by a Doctoral Fellowship from the Canadian Institutes of Health Research (CIHR). J.C.H. and D.C. are fellows of the Leukemia and Lymphoma Society. D.D. is a recipient of the Hitchings-Elion Fellowship of the Burroughs-Wellcome Fund and a Canada Research Chair (tier II) in Proteomics, Bioinformatics and Functional Genomics. This research was supported by grants to J.F.G. from the CIHR and the Ontario Genomics Institute with funds from Genome Canada, to D.D. by the CIHR, to J.E.H. by the National Institutes of Health (NIH), and to S.B. by the NIH.

Author Contributions M.C.K., J.A.K. and M.D. contributed equally to this work.

N.J.K. was responsible for data in Figs 1b (with mass spectrometry help from A.E.) and 2, and Supplementary Figs 3 and 4; M.C.K. in Fig. 1d and Supplementary Fig. 1b-d (with substrate provided by M.O.), Fig. 3c (with D.C. and J.L.), d, and Supplementary Figs 6 and 9; J.A.K. in Fig. 3a, b, and Supplementary Figs 2 and 5; M.D. and D.D. in Fig. 4 and Supplementary Fig. 8a, c; J.F. in Fig. 1a; J.C.H. in Supplementary Fig. 7; S.G. in Supplementary Fig. 1a; N.D. in Supplementary Fig. 8b; and X.S. in Fig. 1c. M.C.K., S.B., J.E.H., D.D., J.F.G. and N.J.K. wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.D. (durocher@mshri.on.ca) or J.F.G. (jack.greenblatt@utoronto.ca).

CORRIGENDUM

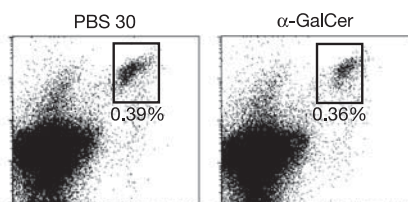
doi:10.1038/nature04475

Exogenous and endogenous glycolipid antigens activate NKT cells during microbial infections

Jochen Mattner, Kristin L. DeBord, Nahed Ismail, Randal D. Goff, Carlos Cantu III, Dapeng Zhou, Pierre Saint-Mezard, Vivien Wang, Ying Gao, Ning Yin, Kasper Hoebe, Olaf Schneewind, David Walker, Bruce Beutler, Luc Teyton, Paul B. Savage & Albert Bendelac

Nature 434, 525–529 (2005)

Figure 3 of this Letter contains an inadvertently duplicated panel: the PBS 30 panel is identical to the α -GalCer panel (top right). The corrected panels are shown here. Our results and conclusions are unaffected by this oversight.



CORRIGENDUM

doi:10.1038/nature04484

Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa A. Bemben, Jan Berka, Michael S. Braverman, Yi-Ju Chen, Zhoutao Chen, Scott B. Dewell, Alex de Winter, James Drake, Lei Du, Joseph M. Fierro, Robin Forte, Xavier V. Gomes, Brian C. Goodwin, Wen He, Scott Helgesen, Chun He Ho, Steve Hutchinson, Gerard P. Irzyk, Szilveszter C. Jando, Maria L. I. Alenquer, Thomas P. Jarvie, Kshama B. Jirage, Jong-Bum Kim, James R. Knight, Janna R. Lanza, John H. Leamon, William L. Lee, Steven M. Lefkowitz, Ming Lei, Jing Li, Kenton L. Lohman, Hong Lu, Vinod B. Makhijani, Keith E. McDade, Michael P. McKenna, Eugene W. Myers, Elizabeth Nickerson, John R. Nobile, Ramona Plant, Bernard P. Puc, Michael Reifler, Michael T. Ronan, George T. Roth, Gary J. Sarkis, Jan Fredrik Simons, John W. Simpson, Maithreya Srinivasan, Karrie R. Tartaro, Alexander Tomasz, Kari A. Vogt, Greg A. Volkmer, Shally H. Wang, Yong Wang, Michael P. Weiner, David A. Willoughby, Pengguang Yu, Richard F. Begley & Jonathan M. Rothberg

Nature 437, 376–380 (2005)

The following were omitted from the original author listing: Alex de Winter, James Drake, Robin Forte, Steve Hutchinson, William L. Lee, Michael Reifler and David A. Willoughby. These names are included in the revised authorship shown here and either were or are at 454 Life Sciences Corporation, Branford, Connecticut 06405, USA.

CORRIGENDUM

doi:10.1038/nature04572

Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*

William C. Nierman, Arnab Pain, Michael J. Anderson, Jennifer R. Wortman, H. Stanley Kim, Javier Arroyo, Matthew Berriman, Keietsu Abe, David B. Archer, Clara Bermejo, Joan Bennett, Paul Bowyer, Dan Chen, Matthew Collins, Richard Coulsen, Robert Davies, Paul S. Dyer, Mark Farman, Nadia Fedorova, Natalie Fedorova, Tamara V. Feldblyum, Reinhard Fischer, Nigel Fosker, Audrey Fraser, Jose L. García, Maria J. García, Arlette Goble, Gustavo H. Goldman, Katsuya Gomi, Sam Griffith-Jones, Ryan Gwilliam, Brian Haas, Hubert Haas, David Harris, H. Horiuchi, Jiaqi Huang, Sean Humphray, Javier Jiménez, Nancy Keller, Hoda Khouri, Katsuhiko Kitamoto, Tetsuo Kobayashi, Sven Konzack, Resham Kulkarni, Toshitaka Kumagai, Anne Lafon, Jean-Paul Latgé, Weixi Li, Angela Lord, Charles Lu, William H. Majoros, Gregory S. May, Bruce L. Miller, Yasmin Mohamoud, Maria Molina, Michel Monod, Isabelle Mouyna, Stephanie Mulligan, Lee Murphy, Susan O'Neil, Ian Paulsen, Miguel A. Peñalva, Mihaela Pertea, Claire Price, Bethan L. Pritchard, Michael A. Quail, Ester Rabinowitsch, Neil Rawlins, Marie-Adele Rajandream, Utz Reichard, Hubert Renauld, Geoffrey D. Robson, Santiago Rodriguez de Córdoba, Jose M. Rodríguez-Peña, Catherine M. Ronning, Simon Rutter, Steven L. Salzberg, Miguel Sanchez, Juan C. Sánchez-Ferrero, David Saunders, Kathy Seeger, Rob Squares, Steven Squares, Michio Takeuchi, Fredj Tekaia, Geoffrey Turner, Carlos R. Vazquez de Aldana, Janice Weidman, Owen White, John Woodward, Jae-Hyuk Yu, Claire Fraser, James E. Galagan, Kiyoshi Asai, Masayuki Machida, Neil Hall, Bart Barrell & David W. Denning

Nature 438, 1151–1156 (2005)

There are two errors in the author listings for this Letter: the surname of Anne Lafon was misspelt as 'Lafton' and the affiliation of Hiroyuki Horiuchi should have been number 16 (and not 15, as was published).

ERRATUM

doi:10.1038/nature04476

Regulated cell-to-cell variation in a cell-fate decision system

Alejandro Colman-Lerner, Andrew Gordon, Eduard Serra, Tina Chin, Orna Resnekov, Drew Endy, C. Gustavo Pesce & Roger Brent

Nature 437, 699–706 (2005)

In Fig. 1b of this Article, the x axis of the right-hand plot should be labelled ' α -Factor system output in each cell (CFP F.U. $\times 10^6$)' and not 'ACT1 system output in each cell (CFP F.U. $\times 10^5$)'. In addition, Supplementary Fig. S6 was incorrect as originally published and was replaced on 26 January 2006.

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Going global

Ten universities spanning four continents this month established an alliance that aims to go beyond the normal exchange of staff and students. The International Alliance of Research Universities is planning to run joint degree programmes, conferences and research projects as well as pooling resources to share knowledge on the commercialization of research. If successful, the alliance could serve as a model for a more global approach to both higher education and research.

Asia is currently represented by the National University of Singapore, University of Tokyo and Peking University, while the Australian National University recently joined the alliance for Australasia. In Europe, the Universities of Oxford, Cambridge and Copenhagen are joined by Swiss Federal Institute of Technology, and the group is completed by the University of California, Berkeley, and Yale University. The initial research areas for the alliance include health; food and water; energy and the environment; security; and the global movement of people.

The opportunities offered by the alliance should prove useful for young scientists who want to establish

international collaborations early in their career. For example, a student from Peking University could spend time in another country during their graduate studies, then use contacts made abroad to strengthen research projects when they return home.

The ultimate success of the plan may well rest on how easy the alliance makes such mobility and how open such exchanges will be. If topics included in the exchanges are limited to a few areas of research, or focused too much on applied research, then the arrangement will be less useful than if it has a much broader remit.

Whatever the case, it is likely that universities that so far have not signed up will be watching closely to see whether they should join the alliance — or perhaps even establish their own.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

Natureevents: Sille Opstrup (4994)
UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
Germany/Austria/The Netherlands:
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/
The Netherlands:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 75 Varick Street,
 9th Floor, New York,

NY 10013-1917
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

School's in for summer

Don't stack shelves in the precious break between terms, stack up your lab experience. **Hannah Hoag** studies the world of the summer intern.

Every spring millions of university students look for summer work. Some wind up in retail, others in the service industry, but those curious about science careers can find themselves redesigning mirrors for the Hubble Space Telescope, developing drugs or testing components of a particle accelerator. Internships exist in all sectors, from academic and research institutions to government laboratories to industry campuses.

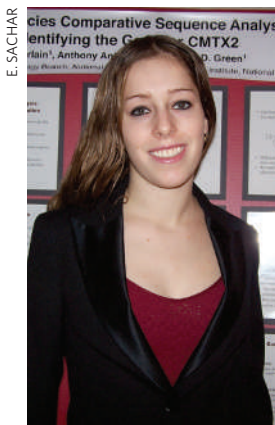
A summer internship is a great way to get experience in a variety of settings. Undergraduates get a chance to try out academia or industry before taking the leap to graduate school. Graduate students can test out a career before signing the contract.

"Internships are an opportunity to try something on for size," says Sara Vancil, a career counsellor and internship coordinator at Princeton University in New Jersey. "An intern will be better equipped to decide on long-term plans and career goals." And these summer positions are far from gofer work. Students make valuable contacts, learn about lab culture or alternate careers, and may even get published.

Summer internships have made Princeton student Amy Wasterlain reconsider her career goals. She has spent two summers in Eric Green's laboratory at the Maryland-based National Human Genome Research Institute, studying enzyme mutations in neurological disorders. Wasterlain has long aimed for medical school, but she is now pondering a joint MD-PhD programme. "I can't practice medicine as an undergrad, only shadow a doctor. But in a research internship you can really get your hands in it. Working on a project that was bench science but also applicable to human disease, I saw how important it was."

Dipping a toe into research

Undergraduates often use the chance to explore research. Most research slots are found in academia or at centres such as the Harvard Stem Cell Institute or the Department of Energy's Fermilab in Batavia, Illinois. These programmes cater to undergraduates in their third and fourth years, although those in their second are also encouraged to apply. "It tells them whether or not they are interested in doing research in astronomy or astrophysics, or whether they are interested in research at all," says Saeqa Vrtilek, director of the summer internship programme at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. Students there may spend their summer analysing data from X-ray satellites or looking for gas emissions from galactic clusters.



Amy Wasterlain: lab work changed her career plans.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Although it may seem that only field biologists get to travel for research, students can get lab experience in interesting locales. The Jackson Laboratory offers students the chance to live in the coastal town of Bar Harbor, Maine. The summer internship programme, which has been around since 1924, takes in about 35 high school and college students — one of them, David Baltimore, went on to win the Nobel prize for medicine in 1975. Places are much sought after, and lab experience can be helpful, says education director Jon Geiger. Marks aren't always the most important criterion: it's more important for applicants to demonstrate that they are self-starters. "We have a pretty high bar for college students," says Geiger. "We basically treat them as beginning graduate students."

The ten positions at the Harvard astrophysics centre and several of those at the Jackson Laboratory make up a fraction of the 4,000–5,000 positions funded by the US National Science Foundation (NSF). For the past few years the NSF has provided about \$50 million to its Research Experiences for Undergraduates programme. This includes a stipend for eight to ten weeks, travel, and room and board. The programme spans all NSF disciplines, from mathematics to polar research, with most projects falling into the fields of biology, chemistry, materials science and engineering.

At NASA, summer students can participate in lab work at different research centres across the country. "One of NASA's missions is to inspire the next generation of explorers," says Dave Rosage, director of the NASA Academy and the NASA Robotics Academy at Goddard Space Flight Center, near Washington DC.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Look and learn: interns can pick up new techniques, make contacts and even become co-authors.

Undergraduates in the summer research programme arrive to conduct high-risk cutting-edge research. Those with some lab experience and aspirations to join the space programme can consider the NASA Academy. Interns work on research problems three days a week, with Thursdays and Fridays set aside for lectures, group project work and field trips to other NASA centres. "People in this programme will hopefully become NASA centre directors," says Rosage.

Graduate possibilities

Although it is less common for graduate students to take on a summer internship, it is not unheard of. A summer post can provide experience they may lack when switching careers, or give a glimpse of a particular path. But graduates should also look at government agencies and industry for summer positions; these could lead to a full-time job.

Government laboratories can offer graduates a multifaceted experience. At the US National Institutes of Health, 1,100 undergraduates and graduates in medicine or dentistry tried out careers in medical sciences for ten to twelve weeks last year.

Felicia Washington, a premedical student at Southern Illinois University with a master's in public health, spent last summer at the National Institute of Child Health and Human Development in Rockville, Maryland. She divided her time between two research projects, one in adolescent problem behaviour and the other in diabetes management. But because Washington was also interested in becoming a clinician, the internship coordinator, Debbie Cohen,

arranged for her to shadow a clinical researcher.

Like academic and government internships, industry internships are highly competitive. The biotechnology research company Genentech, headquartered in San Francisco, California, takes about 5% of applicants and accepts 30 interns for research positions. The posts are often geared towards students with previous laboratory experience, says Monica Poindexter, senior manager of the firm's college programmes. "The philosophy behind the internship programme is really to develop a future talent pipeline for Genentech," she says.

Sizing up the talent

Abbott Laboratories in Illinois also brings graduates and undergraduates to its campus each summer to identify talented individuals, with the intention of hiring them as full-time employees. Students leave with meaningful and transferable work experience, says Christi Lehner, director of college relations.

In addition to research experience, summer internships are full of perks, which students may want to consider when choosing a programme. Lunch-and-learn sessions are popular across the board, offering students a chance to interact with principle investigators, programme directors or vice-presidents.

"You can navigate your way through the organization once you have got through the door as an intern," says Poindexter. "It opens up future positions throughout the company."

For Washington, it was a networking opportunity that allowed her to meet "people who are going to be the movers and shakers in the field, and future colleagues". Students may also have access to study groups that will help them apply to graduate schools.

Most internships provide a window into research culture. At the end of the summer, the majority of students, regardless of their programme, must present the fruits of their labours. It might be a poster for staff and parents or a presentation for board members. For Wasterlain, this was a defining moment: "When you put it on paper you really need to know what is going on. You have to know what you are talking about."

Students often find themselves as co-authors on journal articles. And the summer can have a significant impact on undergraduates returning to start their final-year thesis. Lining up a supervisor is significantly easier when you can go into a meeting and talk about your summer research, list the techniques that you know and give a presentation.

An internship may provide a chance to try out an alternative career. At Jackson Laboratory, science students with an interest in writing can join the lab's public-relations staff for the summer. "They're often not interested in being journalists with a capital J, but they enjoy the teaching part. They enjoy conveying and translating science to the general population," says Geiger. And at Abbott, science interns can learn more about the non-science side of the drug industry by taking a summer position in the quality-assurance or business sectors of the company.

Whether getting a leg-up on graduate applications or testing out a career angle, the experience and relationships gained from an internship can bring young scientists closer to achieving their dream jobs. ■
Hannah Hoag is a freelance writer based in Montreal, Canada.

WEB LINKS

US government agencies
♦ www.science.gov/internships
NSF Research Experiences for Undergraduates
♦ www.nsf.gov/crssprgm/reu/index.jsp

MOVERS

Alan Hall, chairman, cell biology programme, Memorial Sloan-Kettering Cancer Center, New York



2001-06: Director, MRC Cell Biology Unit and Laboratory for Molecular Cell Biology, University College London, UK

1993-2001: Professor of molecular biology, University College London

1981-93: Group leader, Institute of Cancer Research, London

Alan Hall has made his mark in cancer research, but he admits he fell into it almost by accident. He started off in chemistry, studying enzymes as a PhD student at Harvard University. It was the mid-1970s and the molecular biology revolution was under way. He realized that he was more likely to find a long-term, productive future in molecular biology than in enzymology.

"It was clear molecular biology was having an impact on virtually every area of biology," says Hall. And so the stage was set for a career that will take him from Britain to New York this spring, to head the cell biology programme at the Memorial Sloan-Kettering Cancer Center.

Wanting to switch into molecular biology after his PhD, Hall looked for postdoctoral positions in molecular biology labs. His first project at the University of Edinburgh in Scotland didn't work out so well, so he went for another postdoc at the University of Zurich. It was there that Hall mastered the tools and techniques of molecular biology. But he was still looking for an area in which to apply them.

It so happened at the time that Robin Weiss, the director of the Institute of Cancer Research in London, was looking to hire some molecular biologists. Hall was put in touch with him and he landed his first job there in 1981.

It was an exciting time to be doing cancer research. Biologists were beginning to discover the first human cancer genes. Hall and his close collaborator, Chris Marshall, soon became part of this group. In the early 1980s, they identified the third and final of the *Ras* genes, which are involved in 30% of cancers.

After 12 years there, Hall was looking for a change, and so he went to University College London. In his twelfth year there, as he began thinking about his next move (Hall jokes about having a "12-year itch"), he was invited to give a seminar at Memorial Sloan-Kettering. He received an e-mail a few months later about an opportunity to work there. The centre's cell biology programme is growing and the chance to help shape that growth was an incentive for Hall to join the centre.

He says postdocs these days need to be more directed earlier in their careers than he was. He advises young scientists to be careful that they don't always latch on to the latest fad. "Some fads, in retrospect, end up being minor technological breakthroughs," says Hall. Although he switched to molecular biology, it turned out, luckily for him, to be much more than just a passing fad. ■

Corie Lok

BRICKS & MORTAR

Nanoscience factory

By spring, the \$84-million Molecular Foundry on the Lawrence Berkeley National Laboratory (LBNL) campus will allow staff and visiting scientists to begin making new types of nano-scale material — and to help "fact check" existing ones, says Steven Chu, director of the LBNL in Berkeley, California.

Vetting newly created materials is very important in nanotechnology. Many scientists are creating new types of molecule, that, if duplicated, could benefit their colleagues. But few labs are equipped to fully characterize these creations, duplicate the synthesis consistently and watch for fraud.

"There are all sorts of claims in the literature that if you follow this procedure you can make this molecule," says Chu. "Sometimes these processes are less than reliable."

Not only will the foundry have its own nanofabrication, computing and microscopy equipment, but it is also near other unique infrastructure. The Advanced Light Source synchrotron will enable scientists building nanomaterials to quickly obtain three-dimensional images and other structural data. The centre will also benefit from the National Center for Electron Microscopy and the largest US Department of Energy supercomputer available for non-classified research,

both at LBNL. This equipment will help scientists better understand and make predictions on the properties of their novel nanoscale materials.

The Molecular Foundry is recruiting two broad categories of scientists — one to help visiting users tap into the centre's nanofabrication capabilities and another to conduct research on nanostructures. There will also be a group of theoretical scientists to help steer the development of new nanomaterials.

The centre will have a "very strong biology component", says Chu, mirroring the Nobel laureate's interest in biophysics. About a quarter of its \$20-million operating budget will be devoted to biology-related projects. The biological nanostructures labs will be equipped to do mammalian and microbial cell cultures, molecular cloning and genetic and protein engineering.

About 40 projects have already been approved, with two or three people on each. "That's going to grow," Chu says. He adds that the foundry's access to unique computing and imaging resources, as well as its programme of constructing both published molecules and new ones, will attract top people in the field. "It should be nanoheaven." ■

Paul Smaglik is editor of *Naturejobs*.
 ▶ www.foundry.lbl.gov

ALUMNUS JOURNAL

Taking to the air

I am writing this during a holiday in the Swiss Alps. The perspective from a mountaintop is fantastic as I reflect on the past year since I was a *Naturejobs* Graduate Journal writer. It also helps me see the bigger picture on my career.

If you asked me if I would do a PhD again, I would say "yes", even though I'm thinking about leaving academia after I finish this one. Applied biotechnology doesn't seem so hip any more, given the lacklustre funding situation and shrinking number of positions. And I worry about the unclear academic career path.

But I have created an extensive network and gained many valuable skills outside the lab through additional activities. I co-founded and built up the international Young European Biotech Network, where I learned how to build and lead teams and to create relationships with other organizations. I also gained selling and negotiation skills. Moreover, I got to know many people around Europe whom I would never have met at academic conferences.

Now, after talking to many people who have started their own companies, I am thinking about doing the same, and would encourage readers to do so too. Creating your own company has many advantages and an entrepreneur needs many of the same attributes as a scientist: creativity, persistence and an ability to work independently. Together with a couple of friends I am developing a business idea. In a couple of years, you may read about it in this journal... ■

Philipp Angerer is a fourth year PhD student in biotechnology at the Swiss Federal Institute of Technology in Zurich, Switzerland.

Panpsychism proved

Met with stony silence.

Rudy Rucker

"There's a new way for me to find out what you're thinking," said Shirley, sitting down opposite her co-worker Rick in the lab's sunny cafeteria. She looked very excited, very pleased with herself.

"You've hired a private eye?" said Rick. "I promise, Shirley, we'll get together for something one of these days. I've been busy, is all." He seemed uncomfortable at being cornered by her.

"I've invented a new technology," said Shirley. "The mindlink. We can directly experience each other's thoughts. Let's do it now."

"Ah, but then you'd know way too much about me," said Rick, not wanting the conversation to turn serious. "A guy like me, I'm better off as a mystery man."

"The real mystery is why you aren't laid off," said Shirley tartly. "You need friends like me, Rick. And I'm dead serious about the mindlink. I do it with a special quantum jiggle-doo. There will be so many applications."

"Like a way to find out what my boss thinks he asked me to do?"

"Communication, yes. The mindlink will be too expensive to replace the cell phone — at least for now — but it opens up the possibility of reaching the inarticulate, the mentally ill and, yeah, your boss. Emotions in a quandary? Let the mindlink techs debug you!"

"So now I'm curious," said Rick. "Let's see the quantum jiggle-doo."

Shirley held up two glassine envelopes, each holding a tiny pinch of black powder. "I have some friends over in the heavy hardware division, and they've been giving me microgram quantities of entangled pairs of carbon atoms. Each atom in this envelope of mindlink dust is entangled with an atom in this other one. The atom pairs' information is coherent but locally inaccessible — until the atoms get entangled with observer systems."

"And if you and I are the observers, that puts our minds in synch, huh?" said Rick. "Do you plan to snort your black dust off the cafeteria table or what?"

"Putting it on your tongue is fine," said Shirley, sliding one of the envelopes across the tabletop.

"You've tested it before?"

"First I gave it to a couple of monkeys. Bonzo watched me hiding a banana behind a door while Queenie was gone, and then I gave the dust to Bonzo and Queenie, and Queenie knew right away where the banana was."

"I tried it with a catatonic person too. She and I swallowed mindlink dust together and I was able to single out the specific thought patterns tormenting her. I walked her through the steps in slow motion. It really helped her."

"You were able to get medical approval for that?" said Rick, looking dubious.

"No, I just did it. I hate red tape. And

you what," he said finally. "If we're gonna do a proper test, we shouldn't be sitting here face to face. People can read so much from each other's expressions." He gestured towards the boulder-studded lawn outside the cafeteria doors. "I'll go sit down where you can't see me."

"Good idea," said Shirley. "And then pour the carbon into your hand and lick it up. It tastes like burnt toast."

Shirley smiled, watching Rick walk across the cafeteria. He was so cute and nice. If only he'd ask her out. Well, with any luck, while they were linked, she could reach into his mind and implant an obsessive loop centring around her. That was the real reason she'd chosen Rick as her partner for this mindlink session, which was, if the truth be told, her tenth peer-to-peer test.

She dumped the black dust into her hand and licked. Her theory and her tests showed that the mindlink effect always began in the first second after ingestion — there was no need to wait for the body's metabolism to transport the carbon to the brain. This in itself was a surprising result, indicating that a person's mind was somehow distributed throughout the body, rather than sealed up inside the skull.

She closed her eyes and reached out for Rick. She'd enchant him and they'd become lovers. But, dammit, the mind at the other end of the link wasn't Rick's. No, the mind she'd linked to was inhuman: dense, taciturn, crystalline, serene, beautiful...

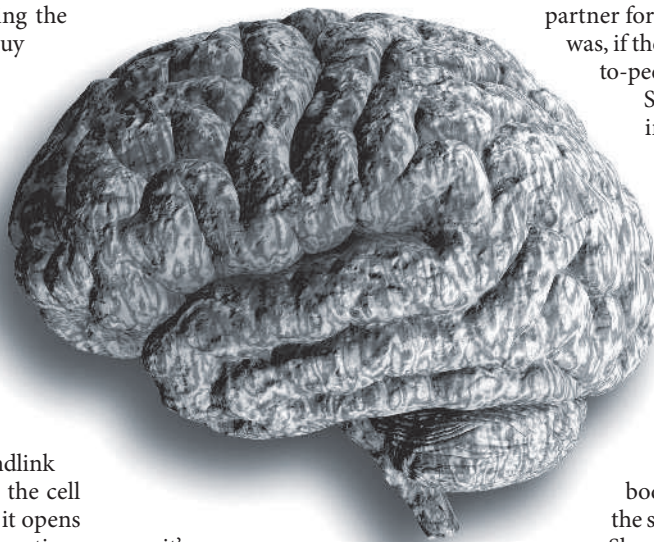
"Having fun yet?" It was Rick, standing across the table, not looking all that friendly.

"What..." began Shirley.

"I dumped your powder on a boulder. You're too weird for me. I gotta go."

Shirley walked slowly out of the patio doors to look at the friendly grey lump of granite. How nice to know that a rock had a mind. The world was cosier than she'd ever realized. She'd be OK without Rick. She had friends everywhere. ■

Rudy Rucker's most recent non-fiction book is about the meaning of computation: *The Lifebox, the Seashell, and the Soul*. See www.rudyruicker.com for further information.



JACEY

now it's time for a peer-to-peer test. With you, Rick. Each of us swallows our mindlink dust and makes notes on what we see in the other's mind."

"You're sure that the dust isn't toxic?" asked Rick, flicking the envelope with a fingernail.

"It's only carbon, Rick. In a peculiar kind of quantum state. Come on, it'll be fun. Our minds will be like websites for each other — we can click links and see what's in the depths."

"Like my drunk-driving arrest, my membership in a doomsday cult and the fact that I fall asleep sucking my thumb every night?"

"You're hiding something behind all those jokes, aren't you, Rick? Don't be scared of me. I can protect you. I can bring you along on my meteoric rise to the top."

Rick studied Shirley for a minute. "Tell